Worcester Polytechnic Institute

Are THP-1 Derived Macrophages a Comparable Model System for Human Dendritic Cells?

A Major Qualifying Project by:

Rachel Brown

WPI Advisor: Elizabeth Ryder

Sponsor: Manuel Garber PhD University of Massachusetts Medical School

April 26th, 2017

Abstract

The immune system is integral to human survival, therefore the ability to understand how the system functions is key to advancing medicine. This research aimed to determine if THP-1 derived macrophages would be a comparable model system for human dendritic primary cells. Chromatin immunoprecipitation and LPS time course RNA-Seq experiments were performed to gather epigenetic and transcriptional data from both cell types. A comparative analysis of enhancer landscape and gene expression profiles in both cell types showed that there is a subset of genes in THP-1 derived macrophages that are comparable to the human dendritic primary cells.

Acknowledgements

Table of Contents

Table of Figures

Table of Tables

Introduction

The human body encounters countless pathogens every day. If these pathogens gain entry to the body they can potentially cause infections and render a person ill. The immune system is thus an extremely important defense mechanism without which human survival is relatively unlikely. The immune system is made up of two types of responses, the innate and adaptive (Parkin and Cohen, 2001). The innate response is the body's first line of defense and includes physical and chemical barriers that are intended to prevent pathogens from entering the body and infecting the host (Alberts et al, 2002). The adaptive immune system is responsible for controlling and eliminating an infection once it has been recognized and is composed of various types of cells that are responsible for recognizing foreign pathogens and destroying them (Alberts et al, 2002).

The focus of this research project is on comparing two types of immune cells that are involved in the innate immune response. The two cell types being examined in this project are THP-1 derived macrophages, an immortal cell line acquired from a leukemia patient, and human dendritic primary cells (HDCs), isolated from donor blood tissue. Macrophages are commonly classified as phagocytic cells that non-discriminately engulf and destroy dead cells, pathogens, and other waste material (Murray and Wynn, 2011). Dendritic cells are characterized for their ability to present antigens, which aids in the communication between the innate and adaptive response systems (Collin et al, 2013).

The intention of this work is to determine if the THP-1 derived macrophages can be used as a model for the HDCs since they are more accessible and easier to genetically modify than HDCs. The ability to model primary cells using immortal cell lines will provide more opportunities to explore and understand more about the immune system without being limited to

the availability and cost of donor samples. To answer the question of whether the THP-1's could be a model system for HDCs, epigenetic and transcriptional data was collected. Epigenetic data was collected by performing a chromatin immunoprecipitation (ChIP) experiment and transcriptional data was collected by performing a time course experiment of response to lipopolysaccharide (LPS), an endotoxin found in the outer membrane of Gram-negative bacteria that elicits strong immune responses.

Background

Immunology

The function of the immune system is to protect and defend the body against foreign pathogens that may cause harm. Every day countless pathogens such as bacteria, viruses, fungi, and parasites attempt to invade and compromise the human body (Parkin and Cohen, 2001). Without any such protection from the immune system, human life would cease to exist. It is the job of this system to be able to prevent, detect, and destroy any foreign pathogen that enters the body before they are able to cause infection. In order to accomplish these tasks, the immune system is composed of two main subsystems, the innate system and the adaptive system (Parkin and Cohen, 2001).

Innate Immune System

The innate immune system is the first line of defense for the body. This subsystem is primarily comprised of physical and chemical barriers such as skin, hair, and mucous membranes as well as lymph nodes, liver tissue, and bile that aim to prevent pathogens from entering or damaging the body. This system is non-specific, such that it responds analogously regardless of the type of pathogen. (Alberts et al, 2002)

The physical barriers like skin and hair prevent pathogens from entering the body. They also contain natural flora that prevent the growth of certain pathogens on the outside of the body. Mucus membranes on the eyes and inside the nasal cavity and mouth contain various enzymes that can kill pathogens and prevent their entry into the system. There are also membranes that contain cilia lining the trachea which prevent pathogen movement. (Alberts et al, 2002)

Chemical barriers such as stomach bile and the liver utilize different chemical properties to damage or destroy pathogens. Stomach bile has a low pH that can kill pathogens before they can cause an infection. The liver produces soluble proteins as part of the complement system that circulate in the blood and become active when they encounter infectious agents acting similarly to antibodies. (Alberts et al, 2002)

Macrophages

Macrophages are phagocytic cells that contribute to the innate immune system. Their main function is to act as a 'janitorial cell' and clear cellular debris throughout the body (Mosser and Edwards, 2009). They are white blood cells that are derived from monocytes in bone marrow, but are also thought to arise during the fetal developmental process as resident macrophages in specific tissues (Epelman et al, 2014).

Macrophages play a role in host defense by remove foreign pathogens such as bacteria, viruses, etc. (Murray and Wynn, 2011). They also can play a role in the inflammatory response, as they aid in wound healing and act as anti-inflammatories. (Murray and Wynn, 2011). The macrophages being examined in this study were derived from THP-1 monocytes and are being examined in response to stimulation by lipopolysaccharide (LPS) an endotoxin known to elicit a strong immune response in animals (Salton and Kim, 1996).

Dendritic Cells

Human dendritic cells (HDCs) are antigen presenting cells that are primarily derived from bone marrow (Collin et al, 2013). They are commonly found circulating in blood, lymphoid, and epithelia tissues (Collin et al, 2013).

HDCs form the bridge between the innate and adaptive immune system. The cells become activated once they encounter pathogen or damage associated molecular patterns (PAMPs) (O'Keeffe et al, 2015). They initiate a primary immune response by taking up antigens that caused their activation and presenting them on the surface of their cell (Clark et al, 2000). Once the antigen is presented, the T-cells can detect the pathogen and activate the adaptive immune response (Clark et al, 2000). The primary HDCs utilized in this study were examined in response to LPS, which as mentioned previously is known to initiate an immune response (Salton and Kim, 1996).

HDCs and macrophages while distinct in their morphology, have been found to have similar functions in regard to cell signaling. A study by Segerer et al reported that the CD68 cell marker was identified on both cell types in a renal biopsy study, suggesting the cell types may overlap in function (Ferenbach and Hughes, 2008). This overlap thus prompts further investigation into the similarities and differences between the cell types as a way to further understand their coordinated action in the innate immune response.

Adaptive Immune System

The adaptive immune response aims to destroy pathogens once they have been recognized by the innate immune system. The response is highly specific and usually long lasting and consists of two classes, antibody and cell-mediated responses.

The antibody response begins when B-cells secrete antibodies that then circulate in the blood stream. The antibodies bind to foreign antigens that were responsible for initiating the antibodies' release. Once bound, the antibody prohibits the pathogen from binding to host cells and marks the pathogen for destruction. (Alberts et al, 2002)

The cell mediated response occurs when T cells react to foreign antigens that are present on the surface of host cells. The T cell either destroys the cell upon which the antigen is present or releases signaling molecules that cause macrophages to destroy the cell. Once a pathogen has been recognized and destroyed, the memory of that pathogen is retained and antibodies against it remain in circulation to prevent repeat infections from occurring. (Alberts et al, 2002)

Cell Lines of Interest

This research focused on two cell types, THP-1 derived macrophages (THP-1 Φ) and human dendritic primary cells (HDCs). Both of these cell types are involved in the innate immune system and aid in protecting and defending the body from foreign pathogens.

THP-1 Cell Line

The THP-1 cell line was isolated from the blood of a one-year old male human Leukemia patient. The isolated monocytes have been used to study the immune response and are also commonly differentiated into macrophages for similar studies. Due to their availability, cost, and genetic manipulability, the THP-1 cell line is an optimal choice for researchers looking to study primary monocytes and macrophages. (Chanput et al, 2014)

The cell line has a number of advantages that make it a good system for study. One main asset is that the cell line is immortal, and because of this the cells can be passaged up to 25 times, which makes them ideal for long-term study. The doubling time of the monocytes is approximately 25 to 50 hours, which is ideal for researchers looking for a faster growth rate. Additionally, the THP-1 cells can be differentiated into macrophage-like cells with phorbol-12-myristate-13-acetate (PMA), which is attractive because ordinarily macrophages would need to

be cultured from isolated peripheral blood mononuclear cells from donor blood samples.

(Chanput et al, 2014)

Primary Human Dendritic Cells

      Primary HDCs are commonly isolated from human peripheral blood mononuclear cells

(PBMC) which come from human donors (Nair et al, 2012). As stated previously, they are

antigen presenting cells that communicate information across the innate immune response to the

adaptive response.

Regulation of Gene Expression

The regulation of gene expression is vital to the success of a cell. By controlling the expression of genes, cells are able to function appropriately and maintain their identity. There are many mechanisms by which regulation takes place. At a macroscopic level, the way in which DNA is packaged into a cell, via condensing into chromatin, provides the most superficial level of regulation. In addition to the physical structure of DNA being packaged as chromatin, chemical modifications on packaging proteins called histones help regulate the accessibility of DNA to transcription factors that aid in gene expression. At the level of the actual DNA sequence, short sequences of nucleotides called enhancers and promoters further regulate how genes turn on, turn off, and produce messenger RNA sequences for the ultimate production of proteins that facilitate cellular functions. (Alberts et al, 2002) The following subsections discuss these topics further and how they relate within the scope of this research project.

Chromatin

The length of the human genome stretched out end to end is approximately 2 meters. The diameter of a nucleus in a human cell is approximately 6 μm. Given this massive difference in size, it is clearly apparent that there needs to be a way in which DNA can be condensed so that it is able to fit inside the nucleus, this condensed packaging of DNA is achieved by the formation of chromatin. (Alberts et al, 2002)
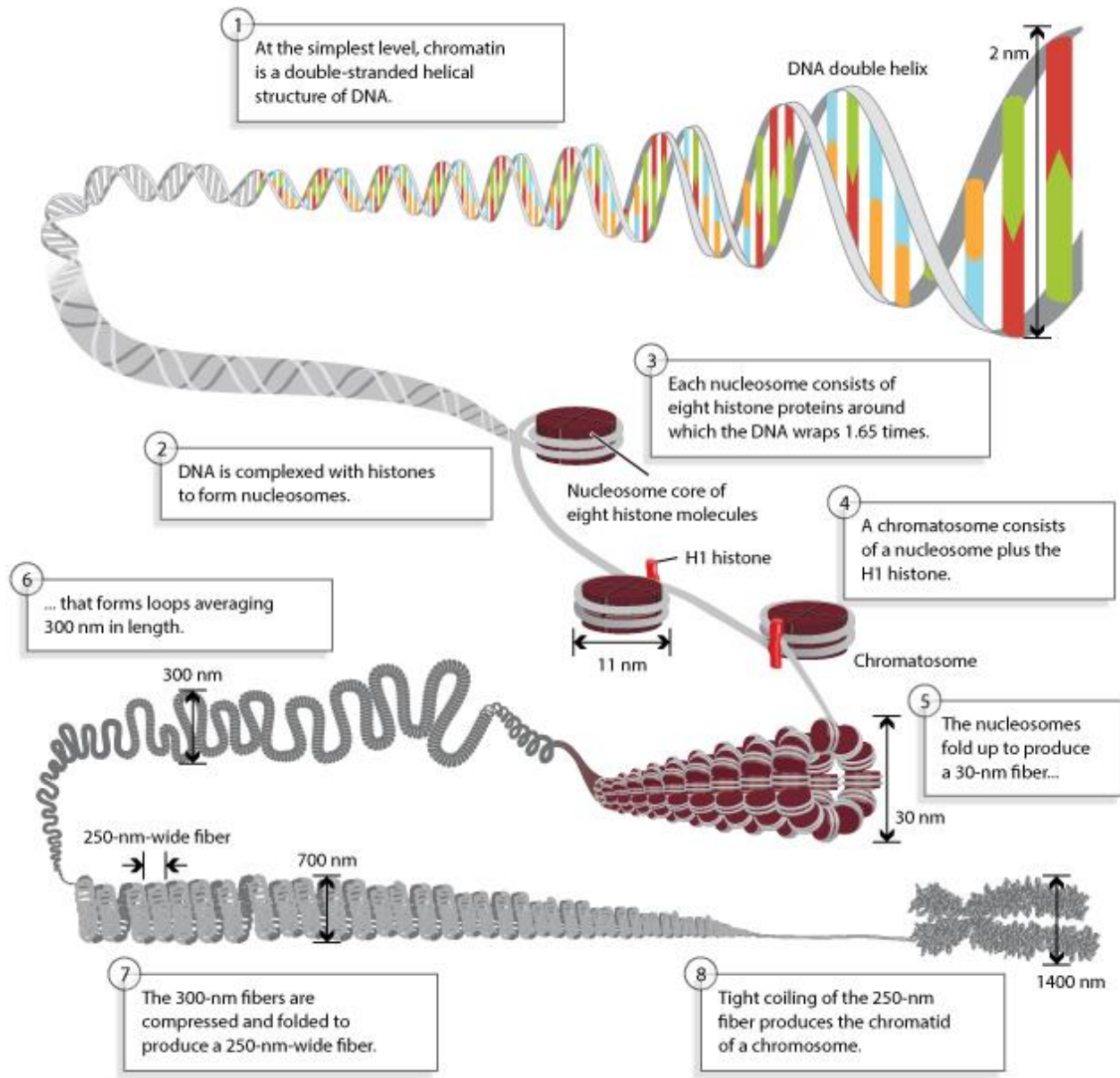
Figure 1. Chromatin has highly complex structure with several levels of organization.
(Benjamin, 2013)

Chromatin can be generally defined as a complex of DNA and proteins that allow the

genome to be tightly packaged into the cell. Chromatin at a basic level starts with the DNA

sequence which is then wrapped around proteins called histones. Histones are protein octamers

that bind to DNA and facilitate the packaging of sequence. The DNA wrapped around the

histone octamers is referred to as a nucleosome. The nucleosomes are then coiled around each other to create a 30nm width fiber that form loops with lengths of around 300nm. These 300nm loops further compact into 250nm fibers that are then coiled further to produce the chromatid of a chromosome. (Annunziato, 2008)

Chromatin is a considerable factor to the regulation of gene expression. If the chromatin is tightly packed, it inhibits expression because the DNA sequence is not being exposed to various transcription factors needed for the transcribing of mRNA. Conversely, if the chromatin is loose and the DNA is exposed, it is easier for transcription factors to bind and for genes to then be expressed. (Annunziato, 2008)

Histones

Histones as mentioned previously, bind to DNA and facilitate the creation of nucleosomes and chromatin. 146 base pairs of DNA wrap around a histone core that is positively charged and is therefore attracted to the negatively charged DNA sequence. The core is an octamer comprised of two sets of the following proteins H2A, H2B, H3, and H4. There is an additional histone protein H1 that functions as the linker protein that aids in the assembly of the nucleosome. (Alberts et al, 2002)

Post-translational Histone Modifications

The eight histone proteins that form the nucleosomes also play a crucial role in regulating gene expression. There are chemical changes that occur at certain amino acid residues in the histone protein sequence called post-translational histone modifications (PTMs) that facilitate changes in chromatin structure. The PTMs can help propagate the formation of looser domains

16

of chromatin also called euchromatin as well as more tightly bound domains of facultative and constitutive heterochromatin. (Bannister and Kouzarides, 2011)

H3K27Ac

A PTM critical to this research project is the acetylation of the 27th lysine on histone 3 (H3K27Ac). This modification is associated with active enhancer regions on the genome. In this research, this marker was used to identify potential active enhancers in unstimulated THP-1 derived macrophages and primary human dendritic cells. (Creyghton et al, 2010)

Enhancers

Enhancers are cis-acting regulatory elements that are composed of short stretches of DNA sequences that serve as binding sites for transcription factors (Shlyueva et al, 2014). It is known that enhancers act at varying distances and orientations from their target promoter (Shlyueva et al, 2014). For this research, active enhancers were predicted by querying the genome for the epigenetic histone mark H3K27ac, which as mentioned above is known to denote active enhancers (Calo and Wysocka, 2013).

Project Goal

The goal of this project was to determine if THP-1 derived macrophages can be used as a comparable model system for human dendritic primary cells (HDCs). To achieve this goal, it was determined that epigenetic and transcriptional data was needed in order to perform a high level comparative analysis of the two cell types. To gather epigenetic data, a ChIP experiment that targeted the histone modification H3K27Ac was suggested since this modification has been associated with active enhancer elements, which would provide insight into how genes are transcriptionally regulated in the cell types (Creyghton et al, 2010). To acquire transcriptional data, a time course study in which cells were exposed to LPS and harvested at various time points was utilized to obtain RNA-Seq data that would provide information about the transcriptional response in both cell types. Since the two cell types are known to be involved in the innate immune response, it was hypothesized that if the THP-1's were a comparable model for HDCs, there would be a subset of genes that contained a similar number of active enhancers and would respond similarly to being exposed to lipopolysaccharide (LPS).

Methods

Chromatin Immunoprecipitation

A chromatin immunoprecipitation (ChIP) experiment was performed for both cell types separately to enrich for the histone modification H3K27Ac, which is used to predict active enhancers. An active enhancer can be described as an enhancer that increases the transcription of a target gene (Heinze et al, 2015). The experiment was performed as described in protocol written by Ronnie Belcher-Gonen published in Nature (Belcher-Gonan, 2013).

LPS Time Course Experiment

Two LPS time course experiments were performed, one for each cell type. Cells were aliquoted into 5 petri dishes labeled 0h, 1h, 2h, 4h, and 6h. A stock solution of LPS was diluted using media to a concentration of 10ng/mL and added to the 1h, 2h, 4h, and 6h dishes. The remaining 0h dish was not treated with LPS. The dishes were scraped, harvested, and a cell count was performed once the dish had been exposed for the appropriate amount of time. After harvest, cells were spun for 5 minutes at 500xg at 4°C. After the first spin, cells were resuspended in PBS. 500μL aliquots were taken from each time point including the 0h time point and pelleted. All aliquots were then placed on ice. 600μL of QIAGEN RLT lysis buffer was added to each aliquot to re-suspend the pellets and then all aliquots were frozen in liquid nitrogen.

Following the time course experiment, the samples were processed using the NuGEN RNA-Seq kit to produce the RNA sequencing data (NuGEn, 2017).

Dolphin Analysis Pipeline

The Dolphin platform developed by the University of Massachusetts Medical School was used to process the raw sequencing data. Dolphin is a computational pipeline designed to handle large datasets that allows users to process raw data (i.e. FASTA files) "end-to-end". Taking in the raw sequence files, users can import their raw data into the system and have it passed through multiple programs such as aligners, peak calling algorithms, etc. to produce usable results. (University of Massachusetts Medical School, 2017).

The raw data collected during the ChIP experiments were processed using the Dolphin ChIP pipeline. The ChIP pipeline aligns reads using Bowtie 2 against the hg19 human genome and utilizes the MACS peak calling algorithm developed by Harvard University to report on sequence enrichment. For this data set the enrichment is for sequences that are associated with the H3K27Ac histone modification that is known to mark active enhancer regions.

The raw data collected from the LPS time course experiments were processed using the RNA-seq pipeline developed by the Garber laboratory. The FASTA files were aligned using the Tophat aligner against the hg19 human genome.

LPS Time Course Filtering Pipeline

After the raw data was preliminarily processed using the Dolphin pipeline, the resulting LPS time course data was filtered based on expression and fold change to isolate genes in both cell types that were similarly expressed and responsive. The filtering was completed using a combination of custom developed KNIME pipelines and custom developed Python scripts. The LPS time course experiment resulted in a data table for each cell type in which each row

corresponded to a single gene and contained the gene name and the transcripts per million (tpm) at each time point sampled (0, 1, 2, 4, and 6 hours post stimulation with LPS).

The data tables for both cell types were processed the following way before they were merged into the master time course table. First, the table was alphabetically sorted by gene name. Next the expression values were normalized by adding 1 to each tpm, this was done to prevent a division by zero error when calculating fold changes. After normalizing the values the $\log_2$ fold change (fc) was calculated from 0h to 1h, 0h to 2h, 0h to 4h, and 0h to 6h. Once the fc values were computed, the genes were filtered using specified tpm and fc threshold values. For each gene if, for any time point the tpm value was not greater than or equal to 25tpm, all tpm values were replaced with zeros. Genes that met or surpassed this tpm threshold were considered to be considerably expressed. For each gene if, for any fc, the $\log_2$ fc value was not greater than or equal to 1, all fc values were replaced with zeros. Genes that met or surpassed this fc threshold were considered to be considerably responsive.

Once the tables for each cell type were filtered, they were merged together into one master time course table. Post-merging, genes were removed from the table if all fc values across both cell types were zeros. Genes were then removed from the table if all tpm values across both cell types were zeros. After the removal of the under-threshold genes, the fc values were removed and the remaining genes were grouped. If the gene had at least one tpm value over 0 for THP-1s and no tpm values over 0 for HDCs the gene was marked as group A. If the gene had at least one tpm value over 0 for HDCs and no tpm values over 0 for THP-1s the gene was marked as group B. If the gene had at least one tpm value over 0 for THP-1s and at least one tpm value over 0 for HDCs the gene was marked as group C. An additional table was created using the

21

grouped genes in which tpm and fc values were reverted back to their original values in order to generate correlation graphs for each cluster.

Genomic Heat Mapping

GENE-E, a "matrix visualization and analysis platform" developed by the Broad Institute was used to generate the genomic heat map (The Broad Institute, 2017). The grouped master time course data table was used as the input matrix. K-means clustering was used to group genes by similar transcriptional response to LPS stimulation.

Epigenetic Data Pipeline

The enrichment data that resulted from processing the ChIP data was output as bed files, in which each row corresponds to an enhancer peak. Enhancer peaks were assigned to the closest gene in which the peak was not located on the promoter site and was upstream of the transcriptional start site. The initial table was processed using a KNIME pipeline to create a table that contained a list of genes along with the number of enhancer peaks associated with each gene, as follows. First, the gene names were extracted from the table. Next, a table was created with each unique gene name and the number of times the gene name was mentioned in the initial table. Since each row in the initial table corresponds to a single enhancer peak, the number of times a gene name is mentioned is equivalent to the number of enhancer peaks for that gene. The table was used to determine the number of enhancers per each cell type, the number of enhancer within each gene group, etc. It is important to note that the enhancers are predicted; to confirm

the enhancer to gene relationship, functional knockouts must be performed to validate the predictions.

Statistics

All statistical analyses were performed using the GraphPad Prism 7 software package. The D'Agostino-Pearson omnibus test was used to determine if the tpm vales of the genes displayed in the genomic heat map followed a normal distribution. The non-parametric Spearman correlation was used to quantitate the correlation of gene clusters identified by generating the genomic heat map. A correlation was considered significant if the p-value for the correlation was less than 0.05 and was considered strong if the r value was greater than 0.5.

Results

   To determine if THP-1 derived macrophages could serve as a comparable model system for HDCs transcriptional and epigenetic data was gathered and a comparative analysis was performed. An LPS time course experiment was performed to gather transcriptional data to determine if there was a subset of genes that were similarly expressed and responsive. A ChIP experiment was done to collect epigenetic data to determine if the THP-1 cell type had a similar number of active enhancers as HDCs. It is important to note that THP-1 derived macrophages, THP-1 Φ, and THP-1's are all terms that are used interchangeably throughout this analysis. Figure 1 below displays the workflow of data generation, processing and analysis.

THP-1 Monocytes   THP-1 Macrophages   HDC Primay Cells

Cell Culture and   Cell Culture and
Harvest   Harvest

ChIP   LPS Time Course

DNA Sequencing   RNA Sequencing

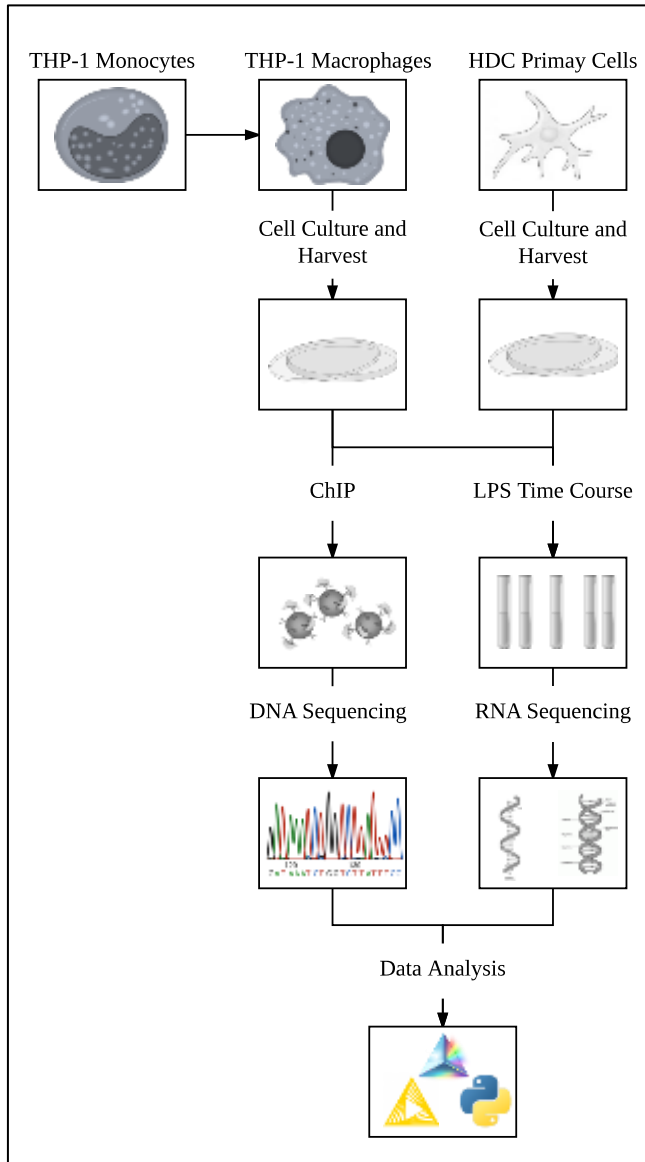Data Analysis

Figure 2. Methods Workflow for Data Generation, Processing, and Analysis

LPS Time Course Results

The goal of the LPS time course experiment was to collect transcriptional data from both cell types before and after they were exposed to the endotoxin LPS. The hypothesis for this experiment was that since the two cell types being investigated -THP-1's and HDCs- are both

25

immune cells with many similar functions, there would likely be a set of genes that were similarly transcriptionally responsive when stimulated with LPS. To test the hypothesis, an LPS time course experiment was performed for each cell type separately. In the experiment cells were sampled prior to stimulation with LPS and also at 1, 2, 4, and 6 hours after stimulation and RNA was extracted and sequenced.

Preliminary Results

Table 1 provides a summary of the preliminary data analysis. There was a total of 25,807 genes annotated for both cell types based on the LPS time course RNA-Seq data. Of these 25,807 genes, 5,997 genes in THP-1's and 5,385 genes in HDCs were identified as expressed at one or more time points (0, 2, 4, or 6 hours after stimulation with LPS). Genes were considered expressed if at any time point the transcripts per million (tpm) was greater than or equal to 25 tpm. There were 3,521 genes identified in THP-1's and 3,028 genes in HDC's that were expressed and also transcriptionally responsive to LPS at any time point. Genes were considered to be responsive if at any time point the $\log_2$ fold change (fc) was greater than or equal to an absolute value of 1. From the preliminary analysis, it can be seen that the THP-1's have more expressed and responsive genes than the HDC's. However, the proportions of expressed and expressed and responsive genes appear to be similar for both at around 20% for expressed genes and just under 60% for expressed and responsive genes.

Table 1. Preliminary Data Analysis

	THP-1	HDC
Total genes annotated	25807	25807
Expressed genes at any time point (tpm > 25)	5997 (23.24%)	5385 (20.87%)
Expressed genes that are also transcriptionally responsive to LPS stimulation at any time point (tpm > 25) (fc > 1)	3521 (58.71%)	3028 (56.23%)

Results in Table 1 suggest that the distribution of expressed and expressed and responsive genes between the two cell types are similar, which provides some preliminary support to using THP-1 derived macrophages as a possible model for HDC primary cells.

Genomic Heat Map

The results from the LPS time course experiment were used to generate the genomic heat map (Figure 2). The input for the heat map was a list of genes and their tpm values for each time point in each cell type. The genes were filtered prior to generating the heat map; any gene within a given cell type that did not meet the thresholds of a tpm value greater than or equal to 25 in at least one time point and an absolute fold change (fc) of greater than or equal to 1 for at least one

fc was "filtered" out by replacing their tpm values with zeros. This filtering resulted in the large

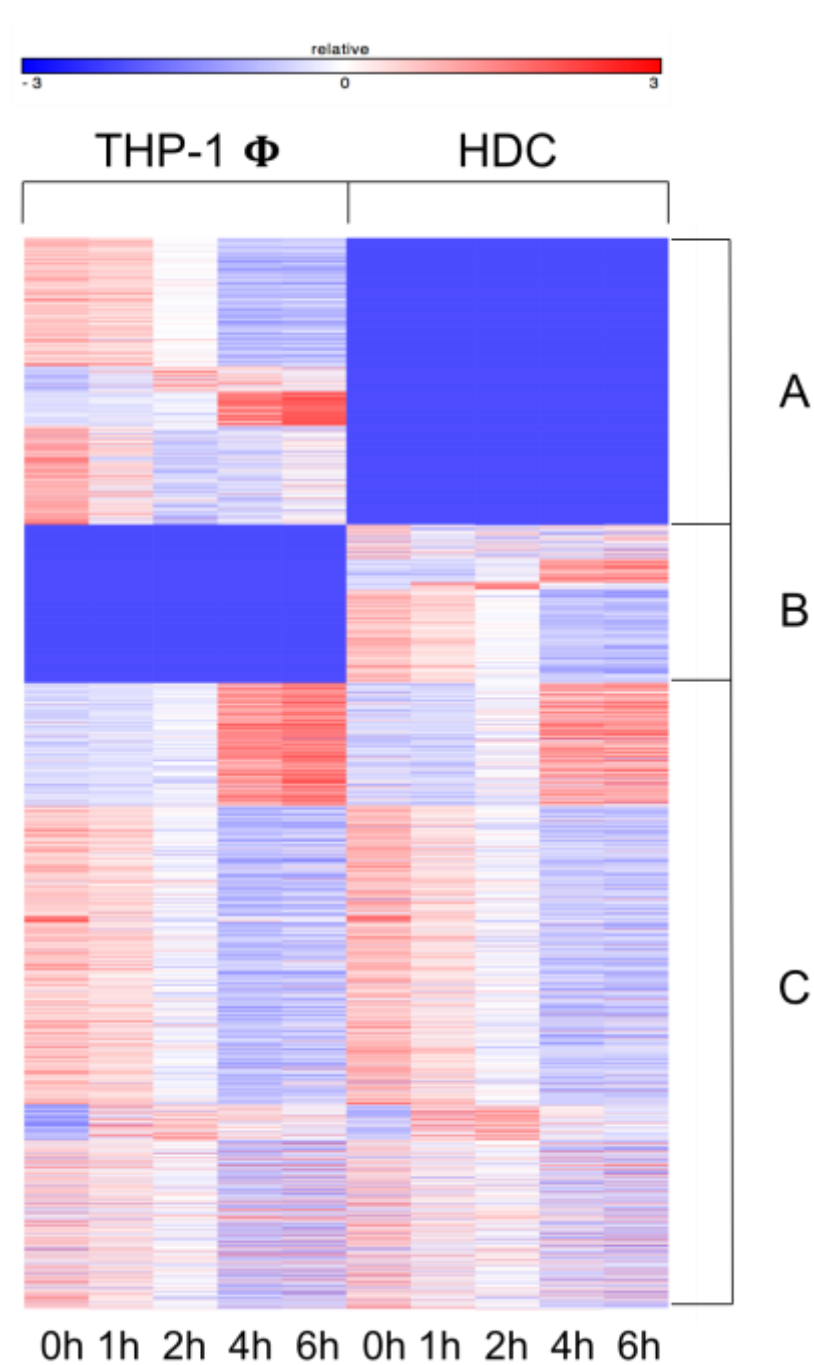blocks of blue genes in each cell type observed in Figure 3.

Figure 3. Genomic Heat Map. THP-1's are represented in the first 5 columns on the left, HDCs
in the remaining columns. Each row represents one gene. Color scale goes from blue to red,

which equates to low to high tpm values. Columns represent tpm values from 0h to 6h. Gene groups A, B, and C were clustered based on expression patterns; see text for details.

The genomic heat map displays the 3 main clusters of genes, referred to as A, B, and C (Figure 2). Group A consists of genes that were expressed and responsive in only THP-1 cells. Group B consists of genes that were expressed and responsive in only HDC cells. Group C consists of genes that were expressed and responsive in both cell types. As stated previously, genes were considered expressed and responsive if at any time point the tpm was greater than or equal to 25 and the fc was greater than or equal to an absolute value of 1.

Group A is composed of 1,106 genes, which accounts for 27% of the total genes included in the heat map, group B contains 611 genes (15%), and group C is comprised of 2,415 genes (58%). From these initial numbers, it appears that the majority of the genes expressed in the LPS time course experiment are genes that were predicted to respond similarly to LPS in both cell types (group C). It can be seen that within group C the majority of genes are downregulated in response to LPS and a small portion are upregulated. The heat map provides additional preliminary support that there are a group of genes that respond similarly to LPS in the two cell types, and suggesting that THP-1's may be able to serve as potential models for HDCs.

Correlational Analysis of Clusters

In order to further quantitate the results from the heat map, the filtered and grouped data table that was used to generate the heat map was utilized to perform a correlation of expression between the two cell types of the genes in each identified cluster. All tpm and fc values that were previously filtered out as 0's were reverted back to their original values for the analysis. To perform this correlation, the absolute maximum fold change for each gene was used as a metric

to compare the transcriptional responses of each gene cluster. Since the data did not pass the

D'Agostino-Pearson omnibus normality test, a non-parametric Spearman correlation test was

performed for each group of genes using GraphPad prism 7. It was hypothesized that genes in

group C would have a strong correlation in expression between the two cell types, since the

cluster was observed to be similarly expressed and responsive in HDCs and THP-1's. Group A

and B were expected to show very weak or non-existent correlations in expression, since genes

in these groups were largely expressed in either one cell type or the other, and not both.

Figure 4 below shows the correlation plot for group A.

Figure 4. Group A Maximum Fold Change Correlation Plot. Each dot represents a single
gene. Maximum absolute fold change was plotted for each gene in each cell type. THP-1's are on
the X-axis and HDCs are on the Y-axis. The Spearman correlation r value and p value are
provided.

It can be seen in figure 4 above that group A gene expression is significantly, but very

weakly correlated between the two cell types, which supports the hypothesis mentioned

previously.

Figure 5 below displays the correlation plot for group B genes, which are largely expressed in HDC cells. The plot was constructed in the same manner as the plot for group A.

Group B Maximum Fold Change Comparison

r = 0.09081
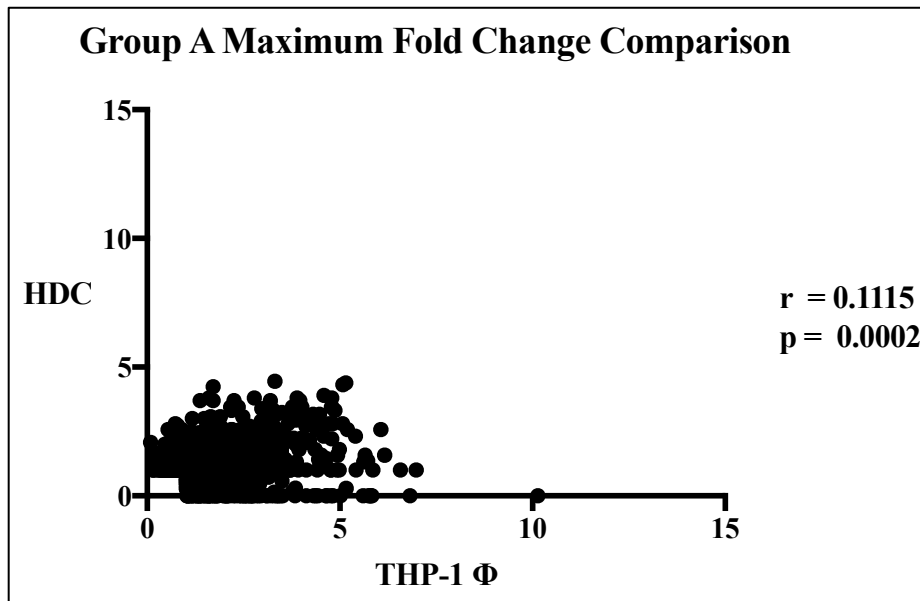p = 0.0248

HDC

THP-1 Φ

Figure 5. Group B Maximum Fold Change Correlation Plot. Each dot represents a single gene. Maximum absolute fold change was plotted for each gene in each cell type. THP-1's are on the X-axis and HDCs are on the Y-axis. The Spearman correlation r value and p value are provided.

From figure 5 above it can be seen that group B has a statistically weak correlation, which supports the hypothesis mentioned above.

Figure 6 below is the correlation plot for group C. Group C genes are expressed and responsive in both THP-1 and HDC cells. The plot was constructed the same way as the previous two plots.
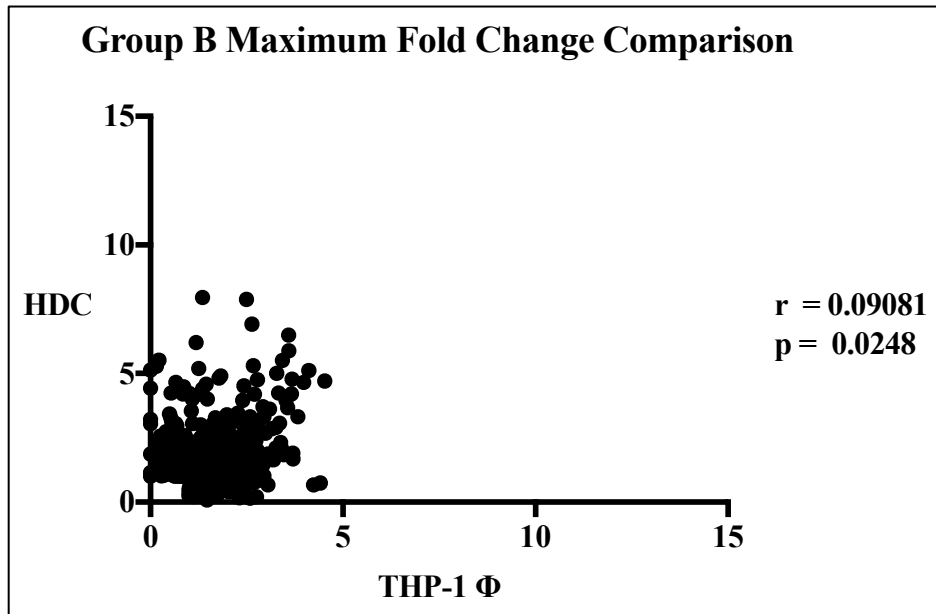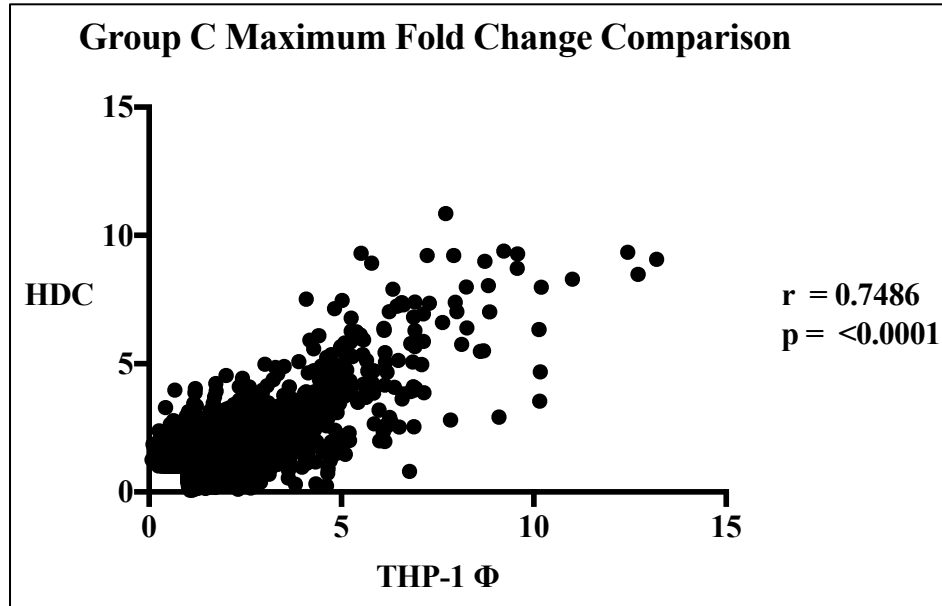
Figure 6. Group C Maximum Fold Change Correlation Plot. Each dot represents a single gene. Maximum absolute fold change was plotted for each gene in each cell type. THP-1's are on the X-axis and HDCs are on the Y-axis. The Spearman correlation r value and p value are provided.

Figure 6 above shows that there does appear to be correlation in group C. The correlation coefficient r was calculated as 0.7486, which means that there is a moderately strong correlation since this value is closer to 1 than to 0. The calculated p value of <0.0001 means that the correlation is statistically significant. Given the results from the correlation, the hypothesis that group C gene expression between cell types would be strongly correlated is supported.

The correlation analyses provide statistical support to the main hypothesis that THP-1's can be used as a model for HDCs provided there is a large subset of genes that are similarly expressed and responsive and have a similar number of enhancers. It was seen that there was a statistically weak correlation observed in groups A and B, which was what was expected, and that there was a statistically moderately strong correlation in group C which suggests that there does appear to be a potential pool of genes that are similarly expressed and responsive.

Chromatin Immunoprecipitation of H3K27Ac Results

The goal of the ChIP experiment was to gather epigenetic data from both cell types that can be used to compare the unstimulated chromatin state of each cell type. The hypothesis for this experiment was that if the cell types were comparable, there would be a similar number and distribution of enhancers among the genes expressed in the LPS experiment. To test this hypothesis, a ChIP experiment was performed using an antibody for the histone modification H3K27Ac, which is a modification associated with active enhancers (Creyghton et al, 2010). It is important to note that the enhancers identified are only predictive, in that their existence and their association with a specific gene must be experimentally validated in order for them to be considered an enhancer and an enhancer associated with a specific gene(s).

Preliminary ChIP Data

The preliminary data shows that the general active enhancer landscape is varied between the two cell types. This was expected as the two cell types do perform different functions and thus must regulate expression of different genes. Figure 7 below displays the total number of active enhancers identified in the ChIP experiment for all genes.
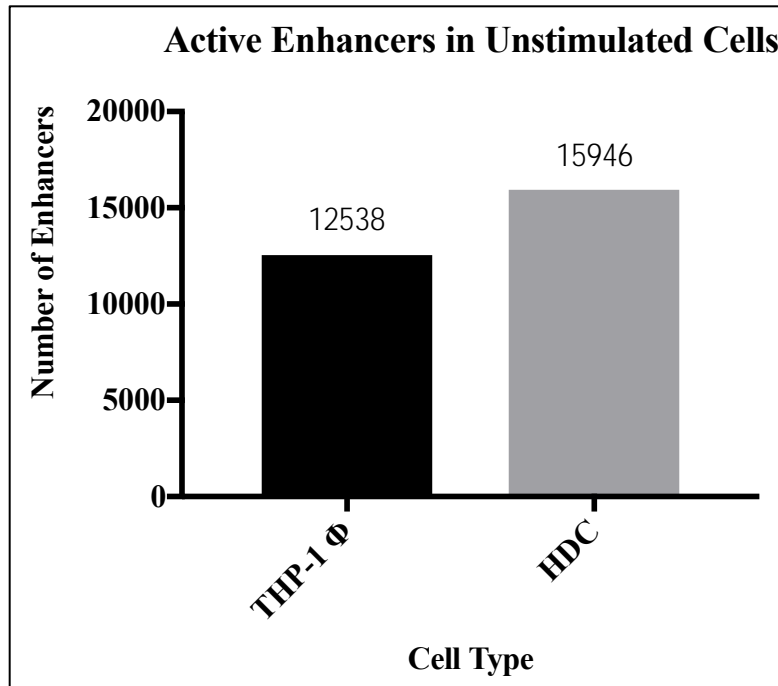
Figure 7. Total Active Enhancers in Each Cell Type. The number of active enhancers identified in all genes for each cell type in the unstimulated state is shown. THP-1's are featured on the left and HDCs on the right. The total for each cell type is provided above the bars.

From figure 7 it can be seen that THP-1's have 12,538 active enhancers and HDCs have 15,946. The HDCs have 3,408 more active enhancers than the THP-1s, which suggests that the two cell types have differences in their overall active enhancer landscape.

Figure 8 below represents the distribution of active promoters identified in all genes for each cell type. A concurrent ChIP experiment was done by members of the lab to collect data on the unstimulated active promoter landscape by using an antibody for histone modification H3K4Me3, which has been associated with active promoter regions (Heintzman et al, 2007). This data was used to correlate active enhancer sequences to the most proximal active promoter. Active enhancer sequences were also eliminated from consideration if they were found to overlap with active promoter regions.
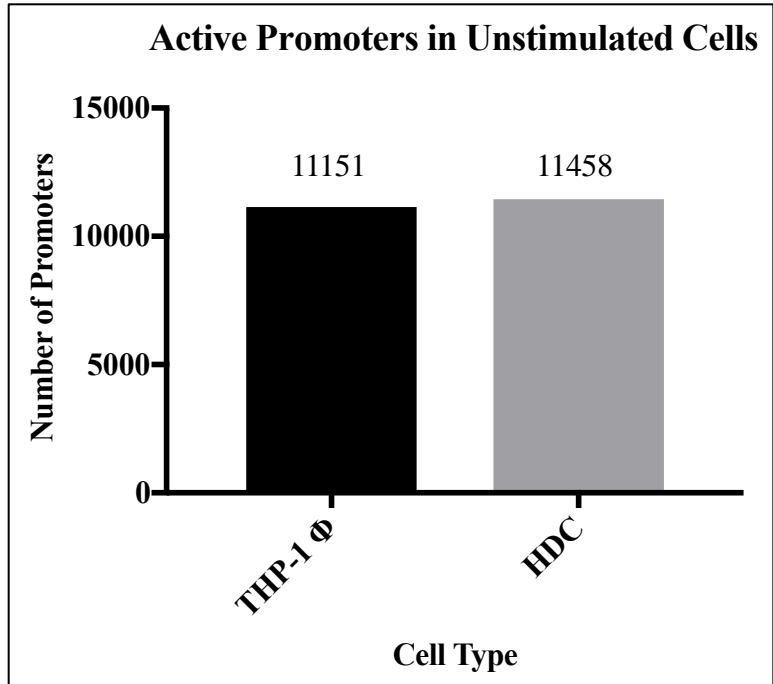
Figure 8. Total Active Promoters in Each Cell Type. The number of active promoters identified in all genes for each cell type in the unstimulated state is shown. THP-1's are featured on the left and HDCs on the right. The total for each cell type is provided above the bars.

It can be seen from figure 8 above that the THP-1 cells have 11,151 active promoters and HDC cells have 11,458. The THP-1's have only 307 fewer active promoters than HDCs.

Active Enhancer Distribution of Gene Clusters

After completing the initial enhancer analysis, the list of potential active enhancers was cross referenced with the clusters of genes identified in the LPS time course experiment. It is important to note that more than one enhancer can be associated with a single gene. The distribution of active enhancers across the three clusters for THP-1's is shown below in Figure 9.
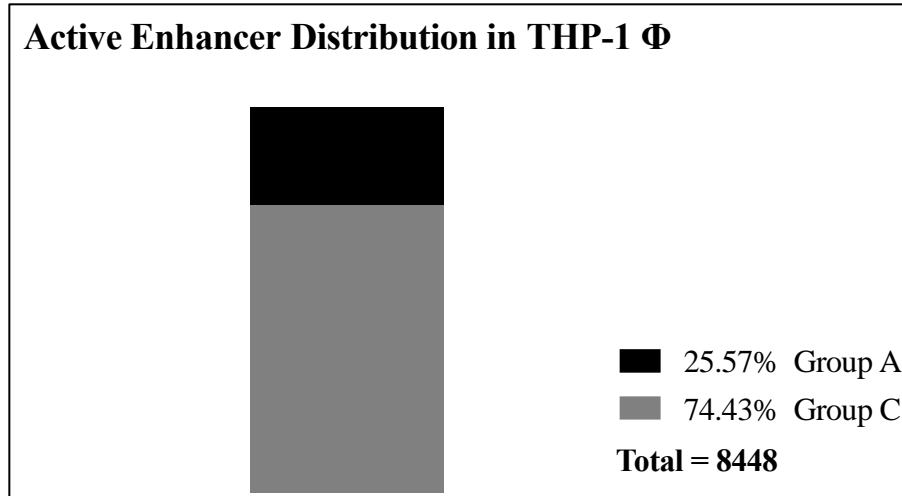
Figure 9. Active Enhancer Distribution in THP-1 Cells. The number of active enhancers within each gene cluster identified from the LPS time course experiment. Group A is the cluster where THP-1's are expressed and responsive and HDCs are not and group C is the cluster where both are expressed and responsive. Note that there were no active enhancers found in group B, the group where HDCs are expressed and responsive and THP-1's are not. Percentages are out of the total active enhancers associated with genes featured in the heat map generated from the LPS time course.

From the figure above it can be seen that 8,448 of the total 11,151 active enhancers identified in THP-1's were associated with 2,225 of the 4,132 genes that were featured in the genomic heat map generated from the LPS time course data. It was found that 25.57% (2160) of the 8,448 active enhancers were associated with 608 of the 1,106 genes in group A and 74.43% (6288) of active enhancers were associated with 1,617 of the 2,415 genes of group C.

Figure 10 below represents the active enhancer distribution across all three gene clusters in HDCs. The stacked bar chart was created in the same manner as figure 9.
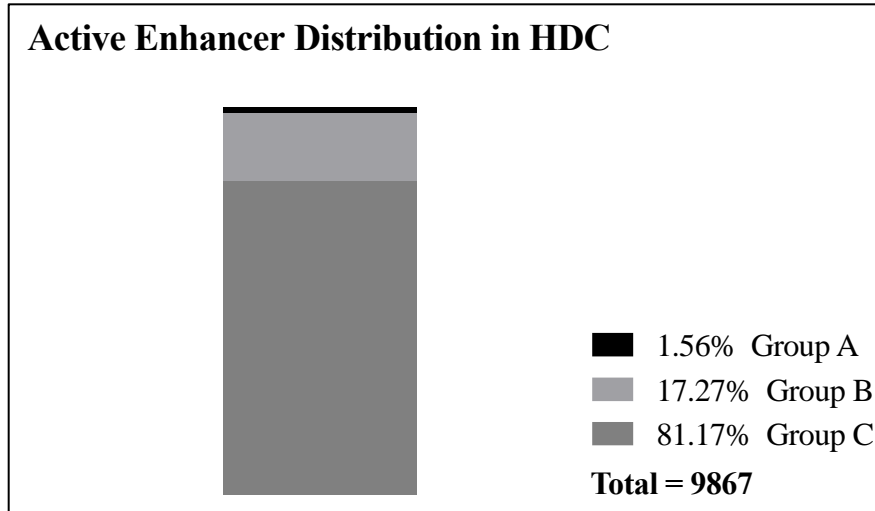
Active Enhancer Distribution in HDC

1.56% Group A
17.27% Group B
81.17% Group C
Total = 9867

Figure 10. Active Enhancer Distribution in HDC Cells. The number of active enhancers within each gene cluster identified from the LPS time course experiment. Group A is the cluster where THP-1's are expressed and responsive and HDCs are not, group B is the cluster where HDCs are expressed and responsive and THP-1's are not, and group C is the cluster where both are expressed and responsive. Percentages are out of the total active enhancers associated with genes featured in the heat map generated from the LPS time course.

From figure 9 it can be seen that 9,867 of the total 11,458 identified HDC active enhancers were associated with 2,162 of the 4,132 genes featured in the genomic heat map. Of the 9,867 active enhancers, 1.56% (154) were associated with 50 of the 1,106 group A genes, 17.27% (1704) were associated with 379 of the 611 group B genes, and 81.17% (8009) were associated with 1,733 of the 2,415 group C genes. Table 2 below shows the percent of genes that have at least one active enhancer in each heat map cluster. This data shows that the H3K27Ac marker does appear to correlate to gene expression in this system. There appears to be a slightly higher percentage of genes with active enhancers in the Group C cluster.

Table 2. Percent of Genes with Active Enhancers in Heat Map Clusters

Cell Type	Percent of group A genes that have at least one enhancer	Percent of group B genes that have at least one enhancer	Percent of group C genes that have at least one enhancer
THP-1	54.97%	0%	66.95%
HDC	0.045%	62.02%	71.75%

Active Enhancer Distribution Across All Genes

After determining the number of active enhancers across the three gene clusters for each cell type, the list of active enhancers was used to generate a Venn diagram that shows the active enhancer distribution across all genes that were found to be considerably expressed. Figure 11 below displays the Venn diagram that was created by cross referencing the list of active enhancers with the genes that were found to have a tpm greater than or equal to 25 at any time point.
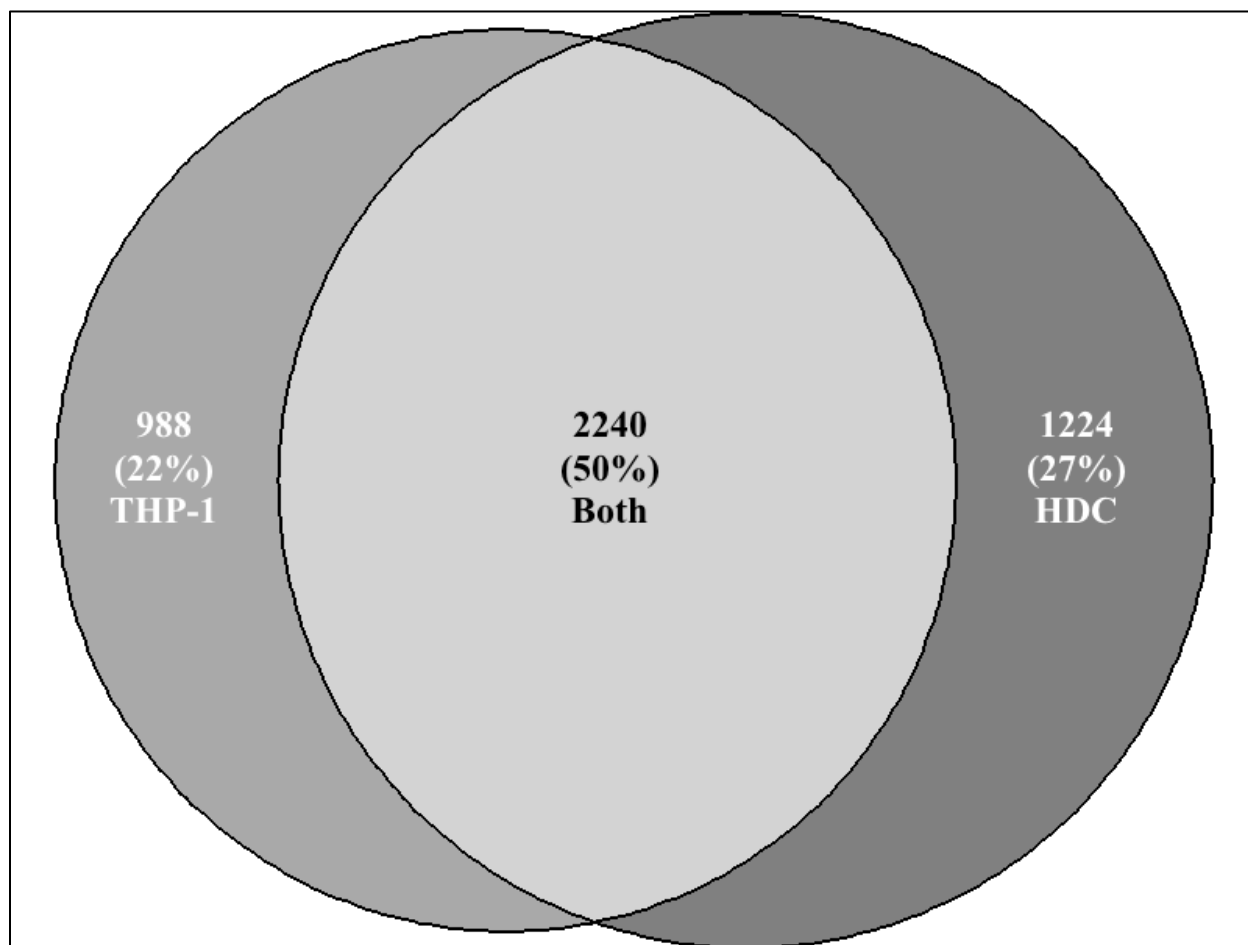
Figure 11. Comparison of Active Enhancer Landscape Across THP-1 and HDC Cell Types.
The Venn diagram shows the number of genes that have enhancers in THP-1's only, in HDCs only, and in both cell types. The size of the circle is proportional to the value inside the circle.

From figure 11 it can be seen that 50% (2240) of considerably expressed genes to which active enhancers were associated were genes that had active enhancers in both cell types, 22% (988) of genes contained enhancers only in THP-1s, and 27% (1224) genes contained enhancers only in HDCs. This data shows that there are a lot of genes that have active enhancers in both cell types. This provides preliminary support to the hypothesis that there are a group of genes that have enhancers in both cell types and thus support that THP-1's can be used as a model for HDCs given their potential similarities in regulation of expression.

Discussion

The goal of this project was to determine if THP-1 derived macrophages could serve as a comparable model system for human dendritic primary cells. After performing both an LPS time course experiment to collect transcriptional information about the cell types and a ChIP experiment to gather data on the active enhancer landscape of both cell types, it can be concluded that there is a large subset of genes in THP-1's that are similarly expressed and responsive and are similarly regulated. This suggests that the cells could serve as potential model for HDCs.

The LPS time course experiment resulted in the identification of three main gene clusters, group A, group B, and group C. Of these three clusters, group C was observed to have genes that had similar transcriptional responses to LPS in both cell types. This suggests that these genes may be comparable to each other since they responded in a similar way to LPS. The next step in this project in terms of further transcriptional comparison of the two cell types would be to perform a gene ontology analysis on group C to determine the functions of those genes. The ontology would provide the distribution of gene function in the cluster and help determine what the genes in the shared cluster are responsible for. It would also provide guidance in terms of choosing genes for future study, as researchers would be able to choose genes that are of a subject area of interest to them such as inflammatory response, adaptive response, etc.

The ChIP experiment results showed that 50% of the genes included in the heat map had enhancers in both cell types. This suggests that these genes are potentially regulated in a similar manner and thus have the potential to be comparable to each other. Future work for the project in terms of epigenetic comparison would be to validate the existence of enhancer sequences by

performing enhancer knockdowns and confirming whether they have any effect on the gene to which they were associated.

Based on all of the preliminary results from this project, there does appear to be support for the use of THP-1's as a comparable model system for HDCs. There was a cluster of genes identified that responded similarly to LPS stimulation and there also appear to be a large portion of genes that contain predictive active enhancers in both cell types. Since these experiments were analyzed using an n of 1, it is important that this work be replicated in the future to validate the results presented. Further work to validate the active enhancers and their relationships to promoters would help determine if indeed the subset of genes in group C in THP-1's have the potential to act as model genes for HDCs. By performing this research, a deeper understanding of the fundamental similarities and differences between THP-1 macrophages and human dendritic primary cells can be used to guide further study of similar and different genes allowing for a greater knowledge and appreciation of how the immune system functions.

References

Alberts B, Johnson A, Lewis J (2002). Molecular Biology of the Cell 4th Edition. New York, NY: Garland Science.

Annunziato, A (2008). DNA Packaging: Nucleosomes and Chromatin. Nature Education. 1(1):26.

Bannister AJ, Kouzarides T (2011). Regulation of chromatin by histone modifications. Cell Research. 21(3):381–395.

Blecher-Gonen R, Barnett-Itzhaki Z, Jaitin D, Amann-Zalcenstein D, Lara-Astiaso D, Amit (2013). High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nature Protocols. 8(3):539–554.

Calo E, Wysocka J (2013). Modification of Enhancer Chromatin: What, How, and Why? Molecular Cell. 49(5):825–837.

Chanput W, Mes JJ, Wichers HJ (2014). THP-1 cell line: An in vitro cell model for immune modulation approach. International Immunopharmacology. 23(1):37–45.

Clark GJ, Angel N, Kato M, López J, Macdonald K, Vuckovic S, Hart DN (2000). The role of dendritic cells in the innate immune system. Microbes and Infection. 2(3):257–272.

Collin M, McGovern N, Haniffa M (2013). Human dendritic cell subsets. *Immunology*. 140(1):22-30.

Creyghton MP, Cheng AW, Welstead GG, et al (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America. 107(50):21931-21936.

Dolphin. University of Massachusetts Medical School. 2015 Dec 10 [accessed 2017 Apr 26]. http://www.umassmed.edu/biocore/introducing-dolphin/

Epelman S, Lavine KJ, Randolph GJ (2014). Origin and Functions of Tissue Macrophages. Immunity. 41(1):21–35.

Ferenbach D, Hughes J (2008). Macrophages and dendritic cells: what is the difference? Kidney International. 74(1):5–7.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Calcar SV, Qu C, Ching KA, et al (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics. 39(3):311–318.

Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. Nature Reviews Molecular Cell Biology. 2015;16(3):144–154.

Mosser DM, Edwards JP (2010). Exploring the full spectrum of macrophage activation. Nature Reviews Immunology. 10(6):460–460.

Murray PJ, Wynn TA (2011). Protective and pathogenic functions of macrophage subsets. Nature Reviews Immunology. 11(11):723–737.

Nair S, Archer GE, Tedder TF (2012). Isolation and Generation of Human Dendritic Cells. Current Protocols in Immunology.

O'Keeffe M, Mok WH, Radford KJ (2015). Human dendritic cell subsets and function in health and disease. Cellular and Molecular Life Sciences. 72(22):4309–4325.

Open for Innovation. KNIME. [accessed 2017 Apr 26]. https://www.knime.org/

Parkin J, Cohen B (2001). An overview of the immune system. The Lancet. 357(9270):1777–1789.

RNA-Seq: NuGEN. 2017 Mar 8 [accessed 2017 Apr 26]. http://www.nugen.com/products/rna-seq

Salton MRJ, Kim K-S (1996). Medical Microbiology 4th Edition. Galveston, TX: University of Texas Medical Branch at Galveston.

Shlyueva D, Stampfel G, Stark A (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nature Reviews Genetics. 15:272–286.