INTRODUCTION

Brown fat is a form of fat that consumes triglycerides to generate heat while white fat stores triglycerides for later use by the body. Brown fat was only recently discovered in adult humans in the supraclavicular region of the neck, while studies of classical brown fat in mice have focussed on the interscapular brown fat in the back, which does not exist in adult humans (Lee, Swarbrick, and Ho, Endocrine Reviews, 2013). Because of the lack of known supraclavicular depots in the mouse, human supraclavicular brown adipose tissue (scBAT) has been proposed to be composed of "beige" or "brite" fat cells, which are cells in white adipose tissue (WAT) that are capable of assuming some of the metabolic characteristics of brown adipose tissue (BAT) in response to chronic cold stimulation (Wu, et al., Cell, 1012). A group at Baylor College of Medicine (Miao-Hsueh Chen's lab) recently discovered BAT depots in the supraclavicular region of the neck in mice while investigating how the Hh signaling pathway regulates BAT development (Nosavanh, et al., PNAS, 2015).

More recently, the Chen lab perfomed RNA-sequencing analysis of this newly identified scBAT depot, the interscapular BAT (iBAT) depot, and two WAT depots. Here, I provide analysis of this RNA-seq data to better determine how the gene expression profiles of the mouse iBAT and scBAT depots compare with those from previously published studies of human scBAT and mouse beige/brite cells. The questions that I attempted to address with this project are 1) does the mouse scBAT gene exression data support the claim that this is a BAT depot? and 2) does the gene expression data indicate that mouse scBAT is composed of classical BAT or of beige/brite cell?

METHODS

The following datasets were used in this analysis: 1) unpublished mouse RNA-seq data for scBAT, iBAT, and subcutaneous WAT (sWAT); 2) human RNA-seq data comparing gene expression in cells derived from scBAT to cells derived from subcutaneous WAT (from Shinoda, et. al., 2015; Supplemental

Table 1); and 3) microarray data of mouse WAT, beige/brite and iBAT cells (from Wu, et al., 2012; data obtained from the Gene Expression Omnibus (GSE39562)).

The mouse RNA-seq data was previously processed to determine significant differences in expression between scBAT and sWAT (SW.vs.VN_DE_excelsorted.csv) and between scBAT and iBAT (VN.vs.IB_DE.excelsorted.csv) using EdgeR. The data in these two files consists of raw mean counts for each sample (counts don't need to be adjusted for gene length because polyA selection used to prepare the RNA-seq libraries, so all of the reads should be from the 3' end, and there should not be duplicate reads for the same transcript) and the fold-changes and p-values of the comparisons. The human RNA-seq data in shinoda.txt consists of FPKM (counts adjusted for gene length) for scBAT and sWAT and the log ratio of scBAT to sWAT; however, only data for genes with two-fold higher expression in scBAT relative sWAT were included in the published table. Three sets of microarray data for brown adipocytes (GSM971717_mouseBAT1.csv, GSM971718_mouseBAT2.csv, and GSM971719_mouseBAT3.csv) and three sets of beige adipocyte data (GSM971716_mousebeige1.csv, GSM971714_mousebeige2.csv, and GSM971698_mousebeige3.csv) were downloaded from the Gene Expression Omnibus (accession number GSE39562). This data consists of the Affymetrix IDs and intensities of the spots on the arrays. I also downloaded a key to these files that provides gene names and Entrez IDs for all of the Affymetrix IDs and saved it as key2wuMicroarrayData.csv.

Matlab was used to generate files and tables containing human and mouse RNA-seq data (CombineRNAseqData.m). To combine human and mouse data, Genbank accession numbers for human orthologs of mouse genes of interest were identified by using the dbOrtho webtool at bioDBnet (https://biodbnet-abcc.ncifcrf.gov/db/dbOrtho.php), and the CombineRNAseqData.m script generated files containing mouse and human RNA-seq data for each gene of interest from the mouse RNA-seq data and all of the corresponding Genbank IDs (i.e., each mouse gene may have been mapped to multiple human genes; this was handled by creating duplicate entries for the mouse genes and generating

separate entries for each human gene). Entrez IDs for mouse homologs for genes in the human dataset were identified using dbOrtho. For these data, the gene was only included if a single mouse gene mapped to the human gene since it was not clear which mouse Entrez ID was the homolog of some of the human Genbank accession numbers. The net result is that only 1785 of the 1950 human genes are included in this set.

Matlab was also used to map Affymetrix IDs to mouse Entrez IDs in the microarray datasets and generate tables and files combining all of the mouse expression data (RNA-seq + microarray; see CombineMouseData.m). Each gene in the mouse RNA-seq dataset was assigned intensity values for the mean of all of the intensities corresponding to that gene in the microarray dataset. This process was repeated for the genes with combined mouse and human RNA-seq data (CombineAllExpressionData.m). Matlab was also used to plot Figures 2B, 3B, and 5. Code for these plots were saved as figure2.m, figure3.m, and figure5.m. PCA analysis was performed after median normalization of each dataset (see figure5.m).

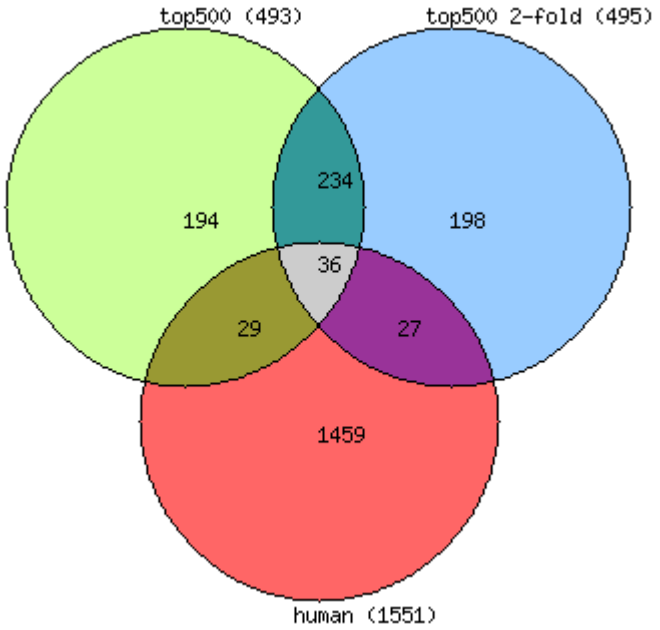
Venn diagrams were generated using the Venny 2.1 webtool (http://bioinfogp.cnb.csic.es/tools/venny/index.html). GO biological process enrichment analysis was performed using DAVID (https://david.ncifcrf.gov). Heirarchical clustering was performed using Multi-experimentViewer 4.9.0 after applying a 0.1% filter cutoff to remove genes that were not expressed in any of the datasets and normalizing each gene to the mean and standard deviation of the corresponding sample. The distance metric used was Pearson Correlation.

RESULTS

Human and mouse RNA-seq datasets are poorly correlated, but the genes common to scBAT in both datasets are enriched for genes involved in brown fat function. Figure 1 shows the overlap between the 500 highest mean counts in scBAT from the mouse RNA-seq dataset with the genes in the human scBAT RNA-seq dataset. The actual number of genes in the set of the top 500 counts was 493

because multiple mean count measures were reported for some genes. Likewise, the set of 1785 mean count entries from human for which mouse orthologs were found corresponded to 1551 unique genes. Only 65 genes overlapped between these two sets; however, GO biological process enrichment analysis revealed that the genes in this set are likely bona fide BAT genes. The top ten GO terms included four processes involving fatty acid and lipid metabolism and "response to cold" (Table 1). One possible explanation for the low degree of overlap between these two datasets is that a significant fraction of the top genes in mouse scBAT may also be highly expressed in mouse WAT. Consistent with this, when I selected the top 500 count measurements from mouse scBAT that were also 2-fold greater than the corresponding counts in WAT, only 270 genes overlapped between the two sets of top 500 counts from mouse. However, only 63 genes overlapped from this set overlapped with the set of genes with 2-fold expression in human scBAT. Again, the genes in this set showed enrichment for fatty acid metabolic biologic processes (Table 1).

Figure 1: Venn diagram of overlap of top 500 counts in mice (top500), the top 500 counts with 2-fold higher expression in scBAT than in WAT in mouse (top500 2-fold), and all genes for which mouse orthologs were identified from the set of human genes with 2-fold higher expression in scBAT than WAT (human).



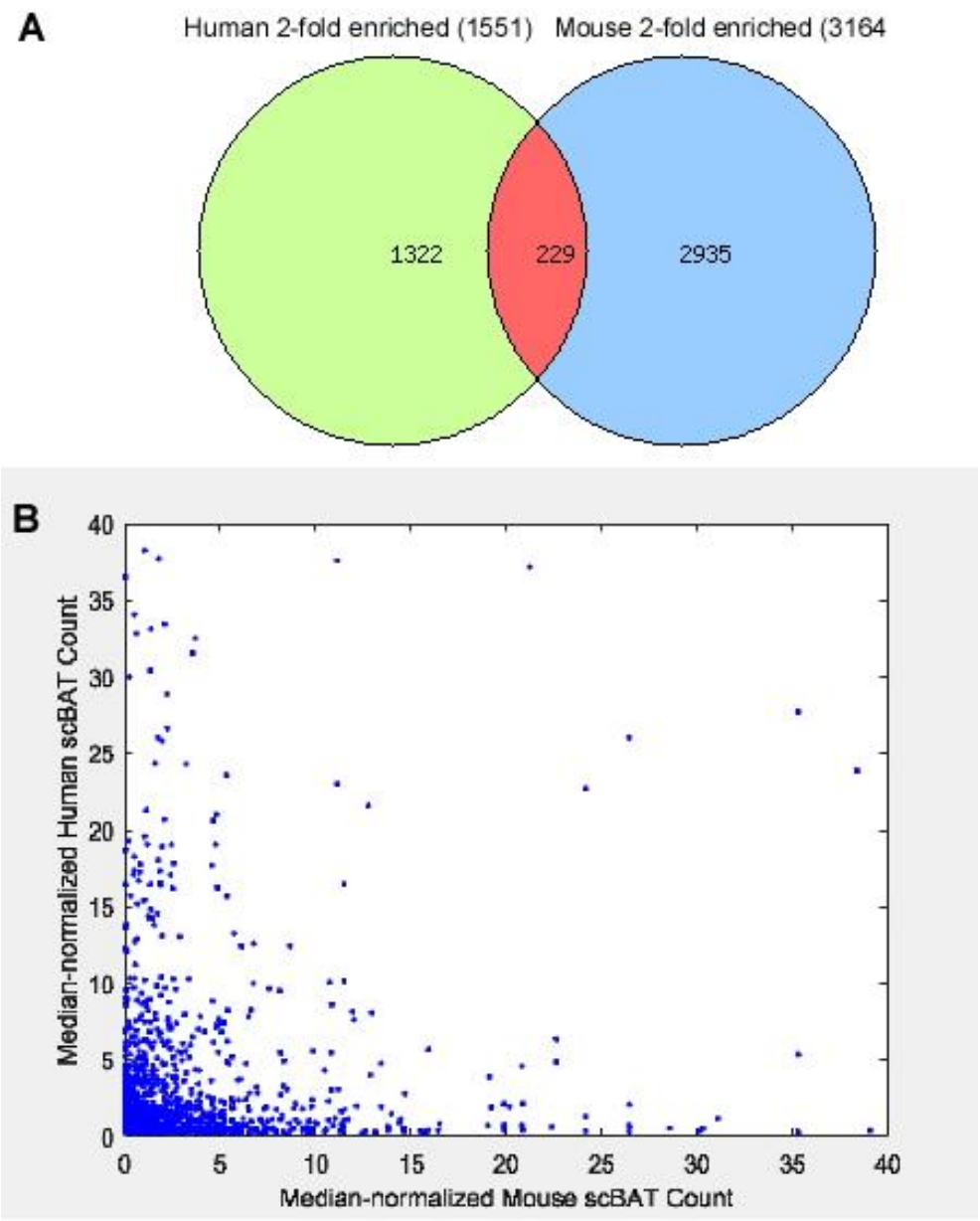


Figure 2: A) Venn diagram of overlap of all genes with 2-fold higher expression in scBAT than WAT in mouse and human. B) Plot showing poor correlation between RNA-seq read counts in scBAT from human with those from mouse. Only genes in the human 2-fold enriched set for part A had counts values in both datasets, and the counts were normalized to the median of the respective set of counts included in the plot.

To further explore how well the two RNA-seq datasets are correlated, I looked at the overlap between all of the genes that showed 2-fold higher expression in mouse scBAT than in mouse WAT (3164 genes) and the genes in the human dataset. As shown in Figure 2A, only 229 genes were found to be 2-fold higher in scBAT than WAT in both datasets. This set also showed enrichment for fatty acid metabolic processes, but more non-specific terms also appeared near the top of list ("cartiledge development"). Plotting the median normalized scBAT counts for all of the genes that were included in

both datasets (where human mouse orthologs mapped to human genes) further indicates that these two RNA-seq datasets are poorly correlated (Fig. 2B).

Because of the poor correlation between expression of the same genes in the two RNA-seq datasets, I investigated if I could mitigate the experimental noise may have affected the data, particularly the mouse RNA-seq data, by focusing on genes that showed statistically significantly higher expression in scBAT than in WAT for mouse. Figure 3A shows that this resulted in slightly better overlap with the human genes that are 2-fold higher in scBAT than in WAT (313 out of 3366), and plots of the log ratios of those 313 genes indicated a slightly improved, but still less than ideal, correlation in expression (Fig. 3B).

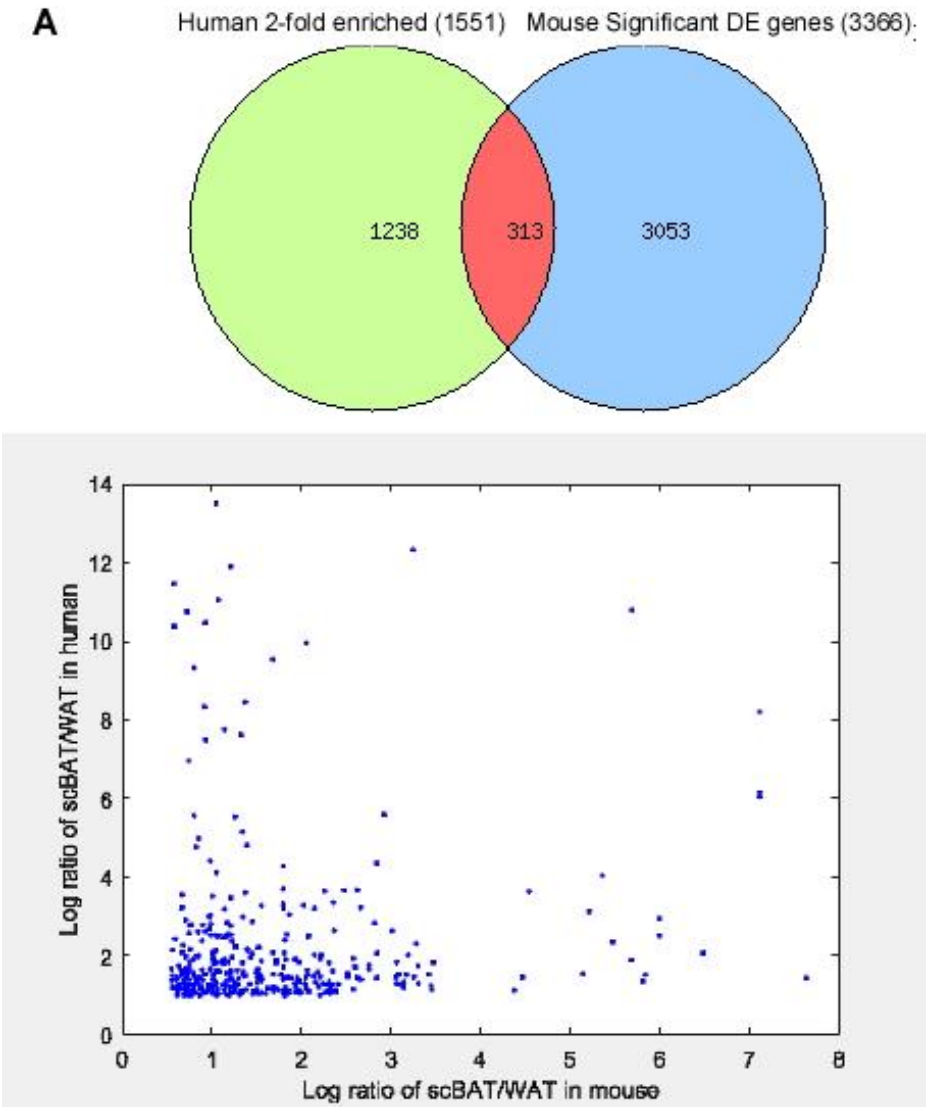


Figure 3: A) Venn diagram of overlap of mouse genes with significantly higher expression in scBAT than in WAT (Significant DE genes) with all genes showing 2-fold higher expression in scBAT than WAT in human. B) Plot of log ratios of scBAT/WAT counts in human vs. mouse for the 313 genes in both sets in A).

Clustering indicates that most of the variance between the three datasets analyzed here can be attributed to 1) experiment-specific factors and 2) species. To determine how closely the gene expression profile of this newly identified mouse scBAT resembles the expression profiles of brown fat vs. white fat, I performed hierarchical clustering on all of the RNA-seq samples and mouse microarray data from three brown fat cell lines (iBAT) and three beige cell lines (Figure 4). Despite using the normalization tool in the Multi-experimentViewer software used for the clustering analysis (mean and standard deviation normalization within each sample), the samples clustered predominantly by study, with the greatest difference being between the human samples and the rest of the datasets. While this meant I was unable to compare mouse scBAT with brown and beige fat expression profiles from other groups, mouse scBAT clustered much more closely to mouse iBAT than mouse WAT, providing additional evidence that this new depot is indeed composed of brown fat tissue.

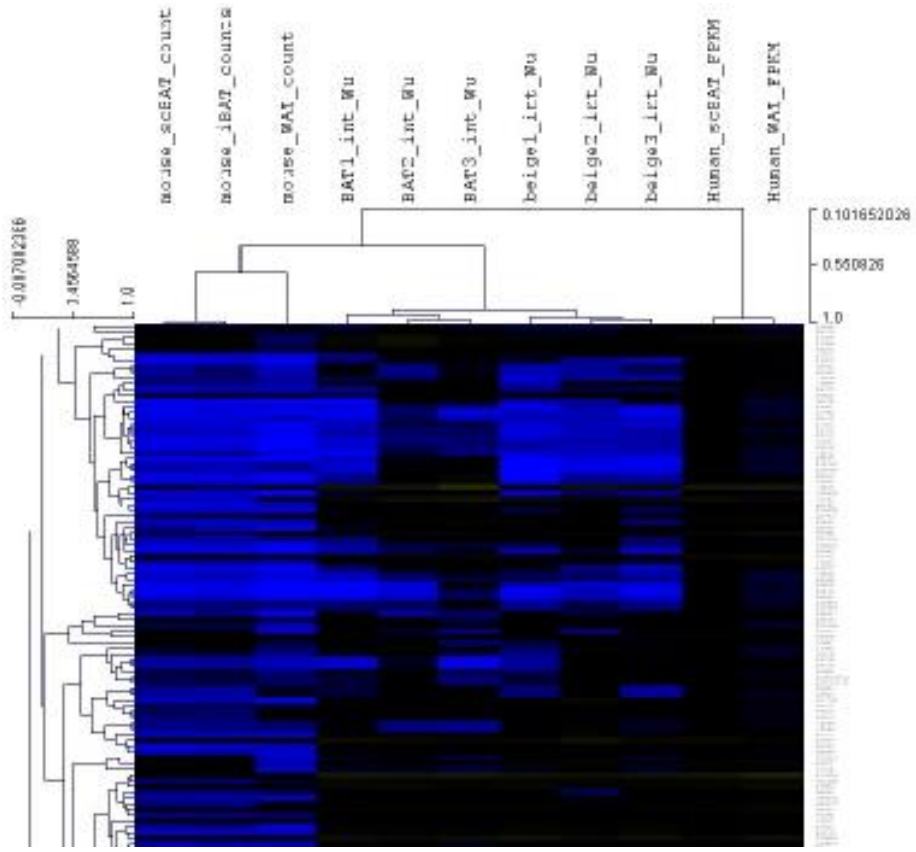


Figure 4: Heirarchical clustering of all datasets. Mouse RNA-seq data for scBAT, iBAT, and WAT, human RNA-seq data for scBAT and WAT, and mouse microarray data for 3 iBAT cell lines and 3 beige cells were clustered using Pearson Correlation as the distance metric.

To better understand the nature of the variance that led to these clustering results, I performed principle component analysis (PCA) on this set on median-normalized data from these datasets (Figure 5). The primary principle component (PC1) separated human WAT from the rest of the samples (note that PCA returned NaN for all of the scores for the human scBAT sample, so it was not included in the plots). PC2, on the other hand, separated our data from the microarray data, although the two WAT RNA-seq samples clustered more together closely along this axis than the mouse WAT RNA-seq clustered with the mouse scBAT and iBAT samples. Interestingly, the two mouse brown fat RNA-seq samples (scBAT and iBAT) clustered very closely with the brown and beige microarray samples, which were closely clustered together in all cases, along both PC1 and PC3.

DISCUSSION

Here I attempted to compare RNA-seq data for a newly discovered mouse suplaclavicular brown adipose tissue depot with RNA-seq data for human scBAT and mouse beige cells. While the experimental noise between datasets made such comparisons inconclusive, compelling evidence that this new depot is actually brown fat was apparent from the analysis. Specifically, the mouse scBAT sample from the Chen lab clustered much more closely to the mouse iBAT sample than it did to mouse WAT. Furthermore, analysis of the genes that did have abundant levels of expression in both the human and mouse scBAT datasets strongly suggests that these genes are involved in fatty acid metabolism, response to cold exposure, and mitochondrial function, all of which are active cellular processes in brown fat. Unfortunately, I was unable to determine if scBAT was more similar to brown fat or beige fat as the beige and brown fat samples clustered more closely together than to any of the other samples I analyzed. PCA did provide some hints as to how we may be able to deal with the over-riding experimental batch effect, however. If PC1 explains most of the variance between the human

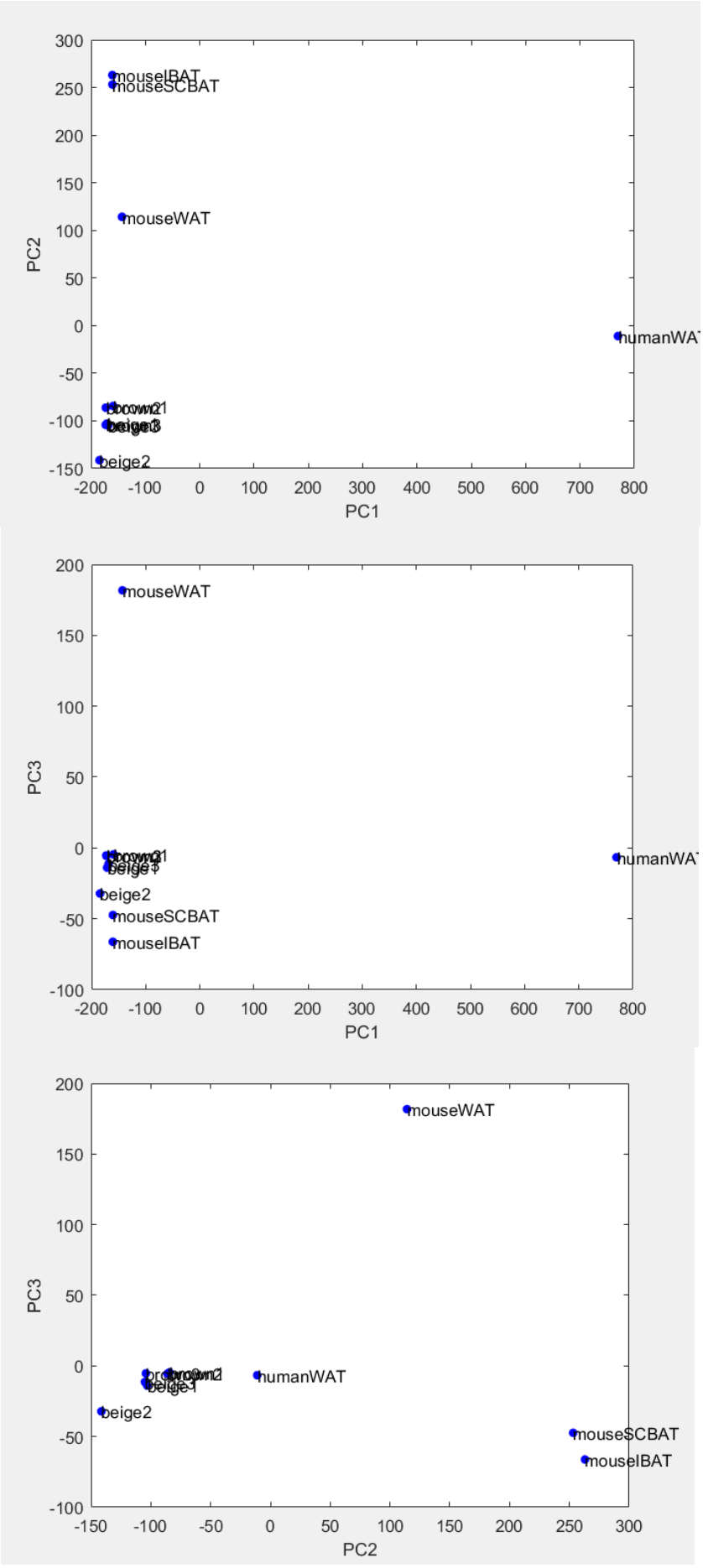


Figure 5: Principle Component Analysis of Chen lab RNA-seq datasets (mouseIBAT, mouseSCBAT, mouseWAT), RNA-seq data from Shinoda et al. (humanWAT), and microarray data from Wu et al. (brown1-3; beige 1-3). Top plot shows principle component (PC) 2 vs. PC1, middle plot shows PC3 vs. PC1, an bottom plot shows PC3 vs. PC2.

expression data and the mouse expression, it may be possible to remove the variance associated with PC1 and achieve better clustering results in the future.

BP Direct Term	Genes	P-value
Figure 1: Overlap of top 500 mouse genes by count with human genes		
GO:0055114~oxidation-reduction process	17	5.09E-10
GO:0008152~metabolic process	14	3.77E-09
GO:0033539~fatty acid beta-oxidation using acyl-CoA dehydrogenase	4	3.38E-05
GO:0006631~fatty acid metabolic process	6	1.79E-04
GO:0050731~positive regulation of peptidyl-tyrosine phosphorylation	5	3.44E-04
GO:0055088~lipid homeostasis	4	3.53E-04
GO:0009409~response to cold	4	3.79E-04
GO:0007568~aging	6	3.83E-04
GO:0006635~fatty acid beta-oxidation	4	4.35E-04
GO:0002042~cell migration involved in sprouting angiogenesis	3	0.001296577
Figure 1: Overlap of top 500 mouse genes showing 2-fold enrichment with human genes		
GO:0008152~metabolic process	14	1.94E-09
GO:0055114~oxidation-reduction process	16	2.20E-09
GO:0005978~glycogen biosynthetic process	4	2.05E-05
GO:0042760~very long-chain fatty acid catabolic process	3	2.09E-04
GO:0009409~response to cold	4	3.27E-04
GO:0006635~fatty acid beta-oxidation	4	3.75E-04
GO:0005980~glycogen catabolic process	3	4.45E-04
GO:0042493~response to drug	7	0.001132806
GO:0006631~fatty acid metabolic process	5	0.001560276
GO:0033539~fatty acid beta-oxidation using acyl-CoA dehydrogenase	3	0.001661221
Figure 2: Overlap of human and mouse genes showing 2-fold enrichment in scBAT relative to WAT		
GO:0055114~oxidation-reduction process	26	1.84E-07
GO:0008152~metabolic process	21	3.21E-07
GO:0009409~response to cold	6	1.04E-04
GO:0006635~fatty acid beta-oxidation	6	1.31E-04
GO:0005980~glycogen catabolic process	4	1.59E-04
GO:0051216~cartilage development	7	3.25E-04
GO:0006631~fatty acid metabolic process	9	3.92E-04
GO:0038084~vascular endothelial growth factor signaling pathway	3	7.45E-04
GO:0005978~glycogen biosynthetic process	4	8.50E-04
GO:0032981~mitochondrial respiratory chain complex I assembly	4	0.001389951
Figure 3: Overlap of human genes with genes that had statistically significant higher expression in mouse scBAT relative to WAT		
GO:0008152~metabolic process	28	4.58E-09
GO:0055114~oxidation-reduction process	32	8.05E-08
GO:0006635~fatty acid beta-oxidation	7	5.77E-05
GO:0032981~mitochondrial respiratory chain complex I assembly	5	2.28E-04
GO:0006629~lipid metabolic process	19	3.33E-04
GO:0005980~glycogen catabolic process	4	4.12E-04
GO:0009409~response to cold	6	4.67E-04
GO:0006631~fatty acid metabolic process	10	7.64E-04
GO:0038084~vascular endothelial growth factor signaling pathway	3	0.001417246
GO:0007005~mitochondrion organization	7	0.001643189

Table 1: Top Ten Gene Ontology Direct Biological Processes Enriched in Sets of Genes Shown in Figures 1 through 3.

REFERENCES

Lee P, Swarbrick MM, Ho KK. Brown adipose tissue in adult humans: a metabolic renaissance. Endocr Rev. 2013. 34:413-38.

Nosavanh L, Yu DH, Jaehnig EJ, Tong Q, Shen L, Chen MH. Cell-autonomous activation of Hedgehog signaling inhibits brown adipose tissue development. Proc Natl Acad Sci U S A. 2015. 21;112:5069-74.

Shinoda K, Luijten IH, Hasegawa Y, Hong H, Sonne SB, Kim M, Xue R, Chondronikola M, Cypess AM, Tseng YH, Nedergaard J, Sidossis LS, Kajimura S. Genetic and functional characterization of clonally derived adult human brown adipocytes. Nat Med. 2015. 21:389-94.

Wu J, Boström P, Sparks LM, Ye L, Choi JH, Giang AH, Khandekar M, Virtanen KA, Nuutila P, Schaart G, Huang K, Tu H, van Marken Lichtenbelt WD, Hoeks J, Enerbäck S, Schrauwen P, Spiegelman BM. Beige adipocytes are a distinct type of thermogenic fat cell in mouse and human. Cell. 2012. 20;150:366-76.