Bios 8060E Project

Amanda Blubaugh

last updated: 10/8/2019

## Project Overview

### Transcriptomic profile differences in atopic dermatitis (AD) canid house dust mite (HDM) models when considering using multifactorial differential expression for sensitization of model dogs

Perform differential expression (DE) of previous data from an AD canine model paper in both a multifactoral DE that mimics the original paper (using R open source versus proprietary software of the paper) and in a classic one factor DE without considering an the additional factor of previous canine sensitization to the HDM model. I would like to see how immunological genes that are significantly up or down-regulated change between the two DE evaluations.

This data set has 12 total canine samples: 6 dogs have been sensitized to the House Dust Mite (HDM) and 6 have not. All dogs are exposed to the HDM and sampled at a 24 hour point after exposure through a patch with HDM crushed paste or a mineral oil patch control. All dogs were also sampled at a 0h, healthy timepoint with no exposure to anything.

The original paper provides a multifactorial differential expression of genes in HDM samples versus 0h control healthy, as well as versus 24h saline, allowing for sensitization of the HDM or not as a factor.

Traditional atopic dermatitis (exzema) canine models for development of human and canine drugs to treat the disease require that all dogs be sensitized to HDM prior to recieving HDM stimulus for modelled skin to AD. canine AD models are essential for drug development in treating AD for humans and canines.

## Goals

Since this paper contains 6 subjects with no previous sensitization, and 6 subjects with sensitization, the goal of this project will be to:

1. Replicate the multifactoral DE of data for all 12 dogs (sensitized or not being the additional factor) at 24h HDM versus 0h untreated no exposure.
2. Do additional individual DE of 6 vs. 6 dogs with no multifactoral consideration for sensitization, since the larger group is divided into two smaller groups that dilineate this factor.

## Hypotheses

It is hypothesized that:

1. Replication of the orginal results will be possible for confirmation of proper manipulation of the data in this study on a multifactoral level and reproducibility to validate the study results.
2. Division of the sample group into smaller, even groups of dogs that have been sensitized or not will provide a DE view that shows differences in DE based on previous sensitization or not, validating whether or not sensitization is required in a HDM canine atopic model.

## Limitations and Justifications

The sample size is quite low for this study, but this is some of the only sample data available for canine AD models to compare sensitized or non- sensitized dogs. We are also not utilizing saline controls of this study, which are still controversial in whether saline itself produces an inflammatory response to the skin.

There is no previous study evaluating the efficacy of first sensitizing dogs to HDM prior to inducing AD models with HDM paste.

## Project Data Source

Samples from a previously published paper by [Schamber et al. (2014)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4199687/) will be used for this project

Schamber et al. Gene Expression in the Skin of Dogs Sensitized to the House Dust Mite (*Dermatophagoides farinae*). G3: Genes, Genomics, Genetics. 2014 Oct; 4(10): 1787-1795.

## Materials and Methods

### Downloading Data from GEOquery (NCBI GEO database)

Prior to working in R studio, I installed [Bioconductor](https://www.bioconductor.org/install/) resources on my local computer through the console, as well as [GEOquery](https://bioconductor.org/packages/release/bioc/html/GEOquery.html) and [limma](https://bioconductor.org/packages/release/bioc/html/limma.html)

if (!requireNamespace("BiocManager", quietly = TRUE))  
 install.packages("BiocManager")  
BiocManager::install()  
  
  
BiocManager::install(c("Biobase","GEOquery","limma"))

### Load libraries for use

library(limma)  
library(Biobase)

## Loading required package: BiocGenerics

## Loading required package: parallel

##   
## Attaching package: 'BiocGenerics'

## The following objects are masked from 'package:parallel':  
##   
## clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,  
## clusterExport, clusterMap, parApply, parCapply, parLapply,  
## parLapplyLB, parRapply, parSapply, parSapplyLB

## The following object is masked from 'package:limma':  
##   
## plotMA

## The following objects are masked from 'package:stats':  
##   
## IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':  
##   
## anyDuplicated, append, as.data.frame, basename, cbind,  
## colnames, dirname, do.call, duplicated, eval, evalq, Filter,  
## Find, get, grep, grepl, intersect, is.unsorted, lapply, Map,  
## mapply, match, mget, order, paste, pmax, pmax.int, pmin,  
## pmin.int, Position, rank, rbind, Reduce, rownames, sapply,  
## setdiff, sort, table, tapply, union, unique, unsplit, which,  
## which.max, which.min

## Welcome to Bioconductor  
##   
## Vignettes contain introductory material; view with  
## 'browseVignettes()'. To cite Bioconductor, see  
## 'citation("Biobase")', and for packages 'citation("pkgname")'.

library(GEOquery)

## Setting options('download.file.method.GEOquery'='auto')

## Setting options('GEOquery.inmemory.gpl'=FALSE)

### Next, download data from GEO datasets

gset <- getGEO("GSE58442", GSEMatrix =TRUE, AnnotGPL=FALSE)

## Found 1 file(s)

## GSE58442\_series\_matrix.txt.gz

## Parsed with column specification:  
## cols(  
## .default = col\_double()  
## )

## See spec(...) for full column specifications.

## File stored at:

## C:\Users\alblu\AppData\Local\Temp\Rtmp0eujTx/GPL18789.soft

if (length(gset) > 1) idx <- grep("GPL18789", attr(gset, "names")) else idx <- 1  
gset <- gset[[idx]]

### Inspect Uncleaned Data

head(gset)

## ExpressionSet (storageMode: lockedEnvironment)  
## assayData: 1 features, 60 samples   
## element names: exprs   
## protocolData: none  
## phenoData  
## sampleNames: GSM1411137 GSM1411138 ... GSM1411196 (60 total)  
## varLabels: title geo\_accession ... treatment:ch1 (42 total)  
## varMetadata: labelDescription  
## featureData  
## featureNames: 1  
## fvarLabels: ID COL ... SEQUENCE (18 total)  
## fvarMetadata: Column Description labelDescription  
## experimentData: use 'experimentData(object)'  
## pubMedIds: 25098772   
## Annotation: GPL18789

head(pData)

##   
## 1 new("standardGeneric", .Data = function (object)   
## 2 standardGeneric("pData"), generic = structure("pData", package = "Biobase"),   
## 3 package = "Biobase", group = list(), valueClass = character(0),   
## 4 signature = "object", default = NULL, skeleton = (function (object)   
## 5 stop("invalid call in method dispatch to 'pData' (no default method)",   
## 6 domain = NA))(object))

## Preliminary Analysis

# make proper column names to match toptable   
fvarLabels(gset) <- make.names(fvarLabels(gset))  
  
# group names for all samples  
gsms <- "013240132401324013240132401324013240132401324013240132401324"  
sml <- c()  
for (i in 1:nchar(gsms)) { sml[i] <- substr(gsms,i,i) }  
  
# log2 transform  
ex <- exprs(gset)  
qx <- as.numeric(quantile(ex, c(0., 0.25, 0.5, 0.75, 0.99, 1.0), na.rm=T))  
LogC <- (qx[5] > 100) ||  
 (qx[6]-qx[1] > 50 && qx[2] > 0) ||  
 (qx[2] > 0 && qx[2] < 1 && qx[4] > 1 && qx[4] < 2)  
if (LogC) { ex[which(ex <= 0)] <- NaN  
exprs(gset) <- log2(ex) }  
  
# set up the data and proceed with analysis  
sml <- paste("G", sml, sep="") # set group names  
fl <- as.factor(sml)  
gset$description <- fl  
design <- model.matrix(~ description + 0, gset)  
colnames(design) <- levels(fl)  
fit <- lmFit(gset, design)  
fit2 <- eBayes(fit)  
topTable(fit2)

## ID COL ROW NAME SPOT\_ID  
## 58100 58100 15 242 CUST\_5667\_PI425858678 CUST\_5667\_PI425858678  
## 39934 39934 71 165 (+)E1A\_r60\_1 (+)E1A\_r60\_1  
## 40727 40727 68 220 (+)E1A\_r60\_1 (+)E1A\_r60\_1  
## 17136 17136 140 170 (+)E1A\_r60\_1 (+)E1A\_r60\_1  
## 35818 35818 83 198 (+)E1A\_r60\_1 (+)E1A\_r60\_1  
## 39232 39232 73 257 (+)E1A\_r60\_1 (+)E1A\_r60\_1  
## 9251 9251 164 196 (+)E1A\_r60\_1 (+)E1A\_r60\_1  
## 36284 36284 82 249 (+)E1A\_r60\_1 (+)E1A\_r60\_1  
## 60894 60894 7 229 (+)E1A\_r60\_1 (+)E1A\_r60\_1  
## 52027 52027 34 251 (+)E1A\_r60\_1 (+)E1A\_r60\_1  
## CONTROL\_TYPE REFSEQ GB\_ACC EntrezGene\_ID GENE\_SYMBOL GENE\_NAME  
## 58100 FALSE   
## 39934 pos   
## 40727 pos   
## 17136 pos   
## 35818 pos   
## 39232 pos   
## 9251 pos   
## 36284 pos   
## 60894 pos   
## 52027 pos   
## UNIGENE\_ID ENSEMBL\_ID ACCESSION\_STRING  
## 58100 ENSCAFT00000042294 ens|ENSCAFT00000042294  
## 39934   
## 40727   
## 17136   
## 35818   
## 39232   
## 9251   
## 36284   
## 60894   
## 52027   
## CHROMOSOMAL\_LOCATION DESCRIPTION GO\_ID  
## 58100 unmapped   
## 39934   
## 40727   
## 17136   
## 35818   
## 39232   
## 9251   
## 36284   
## 60894   
## 52027   
## SEQUENCE  
## 58100 CGGCGGGTGTTGACGCGATGTGATTTCTGCCCAGTGCTCTGAATGTCAAAGTGAAGAAAT  
## 39934   
## 40727   
## 17136   
## 35818   
## 39232   
## 9251   
## 36284   
## 60894   
## 52027   
## G0 G1 G2 G3 G4 AveExpr F  
## 58100 18.89194 18.86503 18.96475 18.84497 18.93218 18.89978 325173.0  
## 39934 18.11555 18.05075 18.42295 18.00749 18.37516 18.19438 271985.5  
## 40727 18.19921 18.12666 18.49375 18.07714 18.44557 18.26847 268295.3  
## 17136 18.11698 18.06180 18.43347 18.01683 18.38568 18.20295 266609.7  
## 35818 18.15314 18.07983 18.44305 18.03145 18.39014 18.21952 264947.1  
## 39232 18.26158 18.18787 18.56994 18.14863 18.51953 18.33751 262678.7  
## 9251 18.17708 18.12538 18.52433 18.09564 18.46371 18.27723 260452.1  
## 36284 18.23556 18.16524 18.54313 18.12610 18.48721 18.31145 260272.3  
## 60894 18.21562 18.19253 18.53691 18.16102 18.50982 18.32318 260157.3  
## 52027 18.26375 18.20822 18.58356 18.17900 18.52945 18.35279 259850.4  
## P.Value adj.P.Val  
## 58100 1.317978e-128 8.300097e-124  
## 39934 2.447089e-126 4.565357e-122  
## 40727 3.648967e-126 4.565357e-122  
## 17136 4.387554e-126 4.565357e-122  
## 35818 5.268483e-126 4.565357e-122  
## 39232 6.774972e-126 4.565357e-122  
## 9251 8.690322e-126 4.565357e-122  
## 36284 8.867580e-126 4.565357e-122  
## 60894 8.982969e-126 4.565357e-122  
## 52027 9.298534e-126 4.565357e-122

## Sample Source Information

1. N\_2847\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411196/> Sample Accession: GSM1411196 ID: 301411196
2. N\_2847\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411195/> Sample Accession: GSM1411195 ID: 301411195
3. N\_2845\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411191/> Sample Accession: GSM1411191 ID: 301411191
4. N\_2845\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411190/> Sample Accession: GSM1411190 ID: 301411190
5. N\_2845\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411187/> Sample Accession: GSM1411187 ID: 301411187
6. N\_2802\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411186/> Sample Accession: GSM1411186 ID: 301411186
7. N\_2802\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411185/> Sample Accession: GSM1411185 ID: 301411185
8. N\_2802\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411182/> Sample Accession: GSM1411182 ID: 301411182
9. N\_2800\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411181/> Sample Accession: GSM1411181 ID: 301411181
10. N\_2800\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411180/> Sample Accession: GSM1411180 ID: 301411180
11. N\_2847\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411192/> Sample Accession: GSM1411192 ID: 301411192
12. N\_2800\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411177/> Sample Accession: GSM1411177 ID: 301411177
13. N\_2716\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411176/> Sample Accession: GSM1411176 ID: 301411176
14. N\_2716\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411175/> Sample Accession: GSM1411175 ID: 301411175
15. N\_2716\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411172/> Sample Accession: GSM1411172 ID: 301411172
16. N\_2704\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411171/> Sample Accession: GSM1411171 ID: 301411171
17. N\_2704\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411170/> Sample Accession: GSM1411170 ID: 301411170
18. N\_2704\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411167/> Sample Accession: GSM1411167 ID: 301411167
19. A\_2855\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411166/> Sample Accession: GSM1411166 ID: 301411166
20. A\_2855\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411165/> Sample Accession: GSM1411165 ID: 301411165
21. A\_2855\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411162/> Sample Accession: GSM1411162 ID: 301411162
22. A\_2853\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411161/> Sample Accession: GSM1411161 ID: 301411161
23. A\_2853\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411160/> Sample Accession: GSM1411160 ID: 301411160
24. A\_2853\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411157/> Sample Accession: GSM1411157 ID: 301411157
25. A\_2841\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411156/> Sample Accession: GSM1411156 ID: 301411156
26. A\_2841\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411155/> Sample Accession: GSM1411155 ID: 301411155
27. A\_2841\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411152/> Sample Accession: GSM1411152 ID: 301411152
28. A\_2840\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411151/> Sample Accession: GSM1411151 ID: 301411151
29. A\_2840\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411150/> Sample Accession: GSM1411150 ID: 301411150
30. A\_2840\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411147/> Sample Accession: GSM1411147 ID: 301411147
31. A\_2839\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411146/> Sample Accession: GSM1411146 ID: 301411146
32. A\_2839\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411145/> Sample Accession: GSM1411145 ID: 301411145
33. A\_2839\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411142/> Sample Accession: GSM1411142 ID: 301411142
34. A\_2838\_24h\_S Organism: Canis lupus familiaris Source name: skin\_treated\_saline\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411141/> Sample Accession: GSM1411141 ID: 301411141
35. A\_2838\_24h\_A Organism: Canis lupus familiaris Source name: skin\_treated\_allergen\_24h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411140/> Sample Accession: GSM1411140 ID: 301411140
36. A\_2838\_0h Organism: Canis lupus familiaris Source name: skin\_untreated\_0h Platform: GPL18789 Series: GSE58442 FTP download: GEO (TXT) <ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM1411nnn/GSM1411137/> Sample Accession: GSM1411137 ID: 301411137