The Correlations of Early Life Nutrition on Lipogenic Genes

Cole Winsor

Statistical Science Biology 7220

Dr. Schneider & Dr. Hurford

Due date (dd/mm/yy): 04/12/23

Submission date (dd/mm/yy): 04/12/23

Editorial Comments (D.Schneider) 08/12/23
Revision submitted

Table of Contents

Introduction

With nearly identical physiology, digestion, and metabolism to humans, the pig is a great model for nutrition/metabolism research [1]. Total parenteral nutrition (TPN) frequently supplies daily nutritional requirements via the bloodstream to people who cannot feed gastrically [2]. A particular case in which TPN is used often is when neonates are born preterm. Many preterm neonates lack the developed physiology for regular feeding, and thus, TPN facilitates the neonates' need for essential nutrients to ensure survival. Since TPN feeding bypasses the digestive tract in a window of high development, there is concern that TPN may induce adverse outcomes that present later in life, such as fatty liver disease [3]. To elucidate the concern of TPN effectively, safely, and ethically on inducing these significant concerns, a piglet model served as the appropriate animal model for the reasons mentioned above. Post birth, yucatan miniature piglets were randomly allocated to one of three diets: sow-fed (SF), TPN control (Ctrl), TPN betaine + creatine (TPN B+C) whereas any piglet <65% the birthweight of the largest littermate was allocated to the TPN intrauterine growth restricted (TPN IUGR) diet group.

IUGR piglets served as the preterm neonates due to their low birth weight.

Since TPN has access to such a critical development window in these patients, stable alternations in gene expression (epigenetic modifications) are possible [3]. Epigenetic modifications can lay the foundations for chronic complications in adulthood, such as dyslipidemia and obesity [4]. The basis for epigenetic modifications is through the methylation of large regions of DNA [5]. A large proportion of whole-body methionine metabolism occurs within the gut [6]. Thus, TPN feeding causes imbalances in the methionine supply, which we believe alters gene expressions and increases obesity and dyslipidemia [6]. To spare methionine via TPN feeding, betaine and creatine are supplemented. Creatine synthesis requires methionine whereas betaine facilitates re-synthesis of methionine [6]. Together, both betaine and creatine limit the use of methionine and increase precursors to produce it. Thus, supplementing betaine and creatine is theorized to reduce methionine imbalances with epigenetic modifications and significantly lower the development of obesity and dyslipidemia in adulthood.

With our foundation set, we can now begin designing the research model to try to answer these questions. To describe how betaine and creatine affect genetic modifications to metabolic

systems that contribute to obesity and dyslipidemia, we have identified specific genes that code enzymes within these pathways. Specifically, fatty acid synthase 1 (FAS-1), acetyl-CoA (ACC-2), stearoyl-CoA desaturase (SCD-1), and are critical to these pathways, thus granting us insight into how epigenetic modifications may be affecting outcomes. We hypothesize that adding betaine and creatine will reduce the expression of these critical genes linked to dyslipidemia and obesity. We must analyze whether there is an increase in mRNA expression versus known standard (housekeeping) genes to answer our hypothesis. A one-way ANOVA will first determine if any of the four means of our diet groups differ. Afterwards, a Dunnett's post-test will elucidate if any group differs from the Ctrl. The SF diet group was significantly different from the Ctrl for FAS-1 only (see Figure 6). Therefore, we will run a multivariant analysis based on diet groups to see any correlations for the genes listed above. Concerning Figure 2, some enzymes (FAS, ACC) are close together in the pathway versus others (DGAT). Therefore, we hypothesize that the enzymes with fewer steps between them will be closely correlated while those with multiple steps will not be. We aim to elucidate whether genes closely linked within the pathway behave similarly in a principal component analysis versus genes on their own.

Methods

Initially, I compared the means of different diet groups for the same gene against each other, which allows the determination of differences of one gene across our diets. However, doing so limits the correlations that may exist due to the effects of multiple genes within a single diet group. For example, FAS-1 and ACC-2 are close together within the pathway (Figure 6); therefore, they may sum to produce a measurable effect within a diet group. For these reasons, we have aligned mRNA expressions within a table so that adjacent columns have matched gene expressions for that diet group. In lay terms, pig #1 mRNA expressions for all genes tested are aligned horizontally and compared to the group (diet) fed to that pig; for this example, it is the control diet. See Table 1. Moreover, our analysis may elucidate whether a gene or cluster of genes influences a weighted measure of our covarying effects, such as our principal components (see Figure 1). With this sorted, we can begin our principal component analysis to determine whether our diet groups affect a set of genes.

Table 1: Example of data setup

	FAS-1	ACC-2	DGAT2-1	Diet	Pig #
mRNA exp	1.11814142	0.66936174	0.37000916	Ctrl	1

Table 1 is an example of how I set up my data within Excel before importing it into R Studio. There are two other genes within the actual table that I mentioned above.

Narrative

A principal component analysis (PCA) includes the correlations among our response variables (gene expressions) which is unfortunately lost when individually analyzing our response variables as we initially completed with an anova. Furthermore, I only looked at differences in means between groups for a single gene, there may be more extensive relationships or correlations between multiple genes and diet groupings which a PCA may elucidate. Therefore, to decide whether various genes are affected by the diet groups for this study, we will run a PCA of five genes specific to the lipogenic pathway.

Initially, I did not have access to the code that allowed for this analysis, so I turned to ChatGPT. Using reports from previous years, I gathered enough to create prompts specific enough for ChatGPT to make code that served as a starting point. I could replicate plots from earlier reports with a few errors through trial and error. ChatGPT helped me troubleshoot these errors and refine my code. However, ChatGPT, as the course instructor put it, is not great at science. It can tell me what the code does but not what the output (plot) represents. I've

attached the code I constructed with ChatGPT and code from a previous students project [7] at the end of this paper for viewing.

To begin, I organized my data into a table within Excel; see Table 1 for an example with the actual data attached separately. Once imported into R studio, I got a summary of which components accounted for the most significant variance (Figure 1).

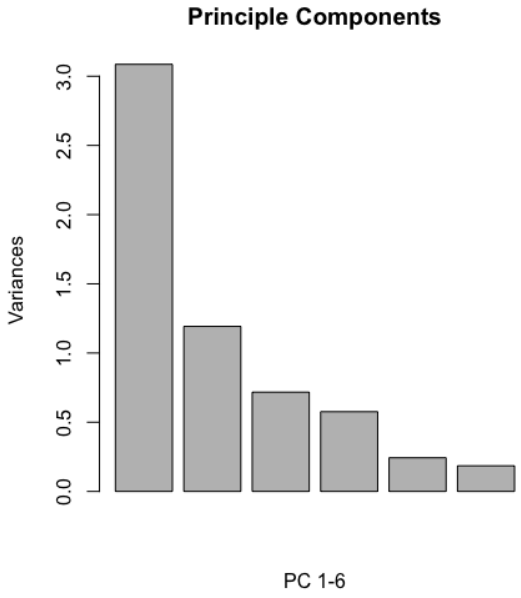

Figure 1: PCA Eigenvalues. *Left to right PC1 to PC6. As shown, PC1 & 2 account for the majority (~70%) of the variance.*

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	1.7567	1.0920	0.8469	0.75900	0.49304	0.43009
Proportion of Variance	0.5143	0.1988	0.1195	0.09601	0.04051	0.03083
Cumulative Proportion	0.5143	0.7131	0.8326	0.92866	0.96917	1.00000

Based on Figure 1 and the proportion explained within the table above, we concluded that PC1 and PC2 account for ~70% (71.3) of the variance. More than 60% variance within two principal components is generally considered sufficient. Thus, we will use these to help answer our hypothesis and goals.

Next, we must examine how the loading vectors (variables), genes and diet, drive PC1 and PC2.

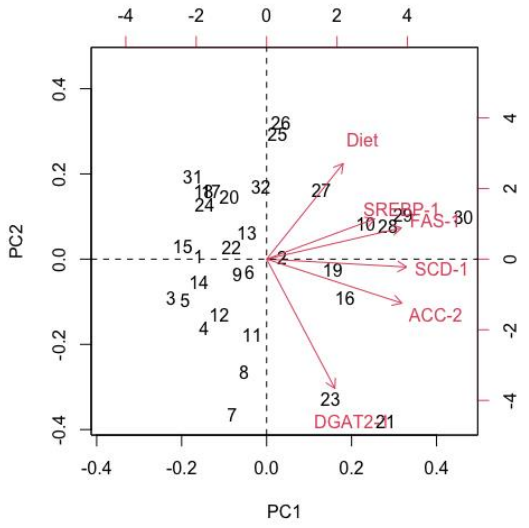

Figure 2: *Genes and Diet plotted against PC1 and PC2.*

Within Figure 2, I'm looking for the longest arrows in the direction of PC1 and PC2. The direction along the y-axis designates influence on PC2, whereas the x-axis represents PC1. We can see that DGAT2-1 appears to have the longest arrow and, besides Diet, is the only variable with a noticeable influence over PC2. DGAT2-1's unique influence over PC2 appears again below when we pull out each gene and its influence over a single PC. Additionally, all variables positively influence PC1, with SREBP-1, FAS-1, SCD-1, and ACC-2 being clustered together; ACC-2 and SCD-1 appear to have the strongest influence. Based on Figure 2, I will say that a value > 0.2 is strongly correlated with the respective PC. Therefore, only DGAT2-1 and Diet do not

strongly connect with PC1, whereas they are the only variables that strongly correlate with PC2.

Unfortunately, I could not get a Diet scores plot that designated my Diets using Molly's code [7]. However, I could do so with the code from my ChatGPT trials. Therefore, we will interpret that plot.

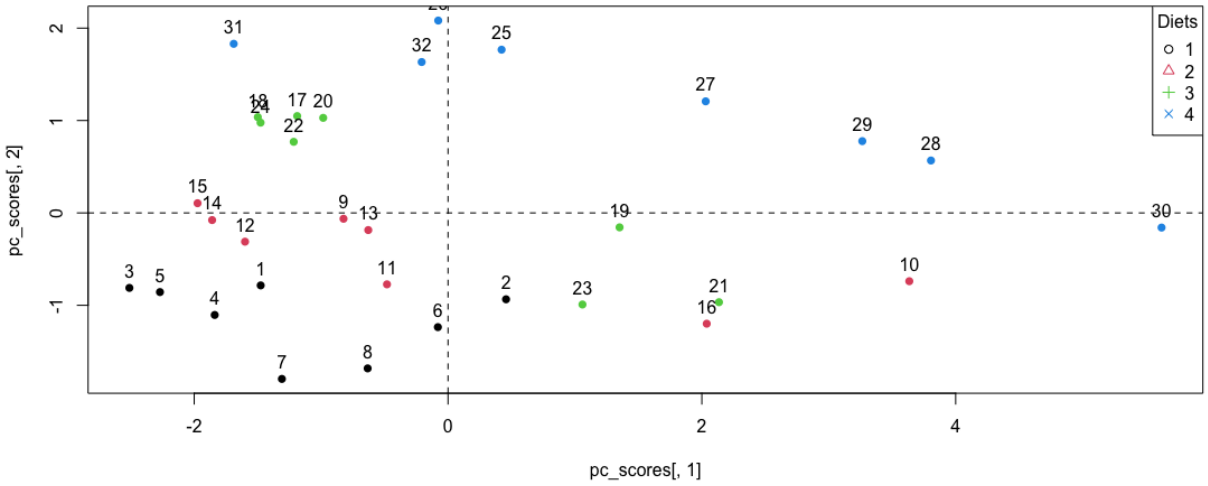

Figure 3: *Site scores vs PC scores.*

Within Figure 3, I'm looking for clusters of similar colours and their correlation with the PCs. To orient ourselves, diet 1: Ctr; diet 2: B+C; diet 3: IUGR; diet 4: SF. Diet 1 (Ctrl) appears to have a strong negative correlation with both PC scores. Diet 2 (B+C) only has a strong negative correlation with PC 1 scores. Diet 3 (IUGR) has a wide range with a cluster of strong positive correlation with PC 2 scores. Finally, Diet 4

(SF) has the strongest correlations with scores and the most comprehensive range.

Finally, we can look individually at each variable (gene and diet) versus the two principal components to see their effects. Based on our biplot (Figure 2) and site plot (Figure 3), we expect DGAT2-1 to be unique within the PC1 plots due to its essentially vertical effect arrow shown above. Additionally, with the remainder of the genes mainly horizontal, we expect them to have weak or no correlations with PC2. With our diets, we expect most of its influence to be against PC2 due to the predominantly vertical arrow; thus, we should expect a positive correlation.

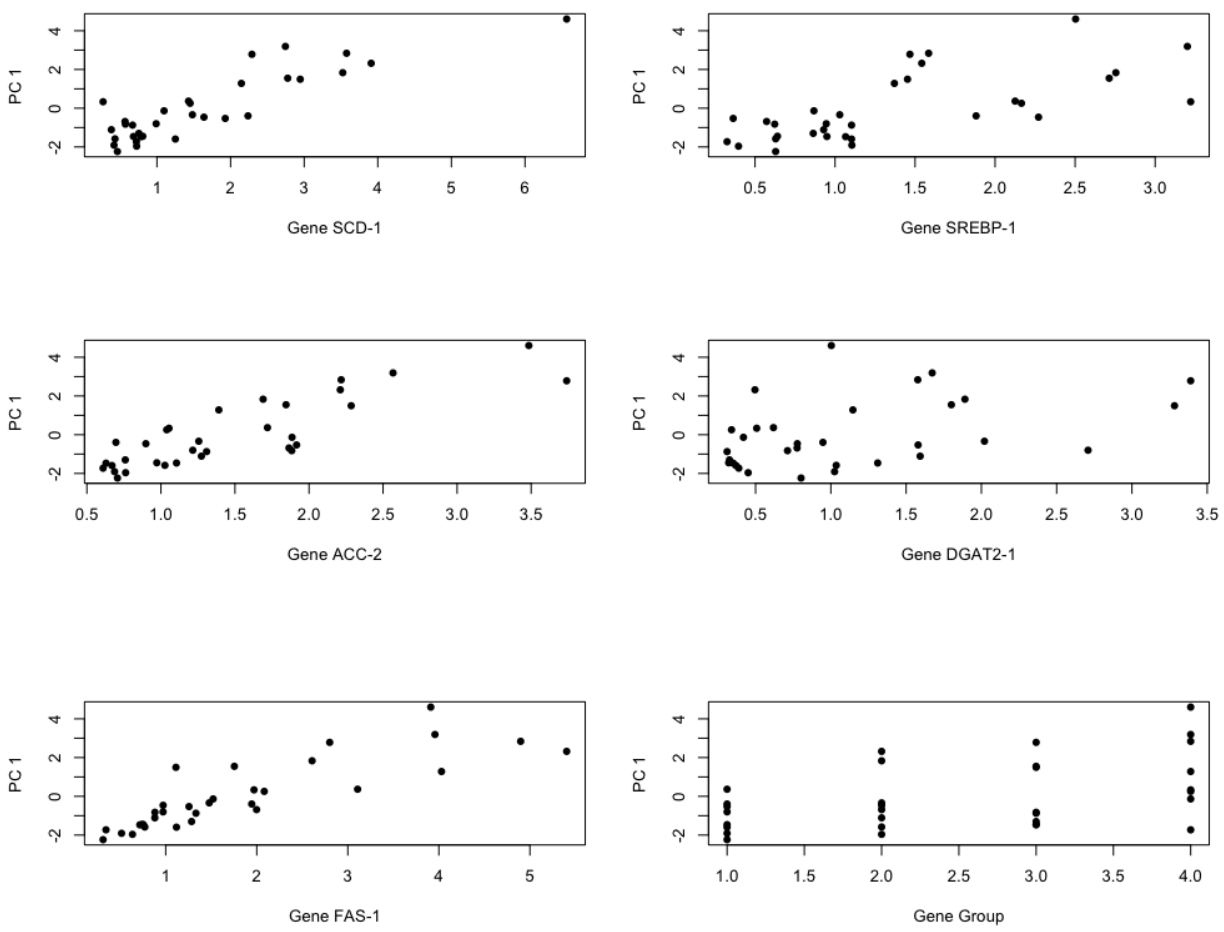

Figure 4: *PC1 versus all variables individually.*

As expected, only DGAT2-1 versus PC1 demonstrates no correlation. All other genes are driving PC1 besides DGAT2-1. There are likely many reasons why DGAT2-1 is different from the rest of the mRNA expressions of lipogenic pathway genes. DGAT2-1 is responsible for converting diacylglycerols into triacylglycerols, the primary storage form for fatty acids within fat tissue. The piglets presented with higher backfat

within the IUGR after feeding; therefore, it could be that DGAT2-1 was at a higher activity than the other genes. It may also be due to the location within the pathway.

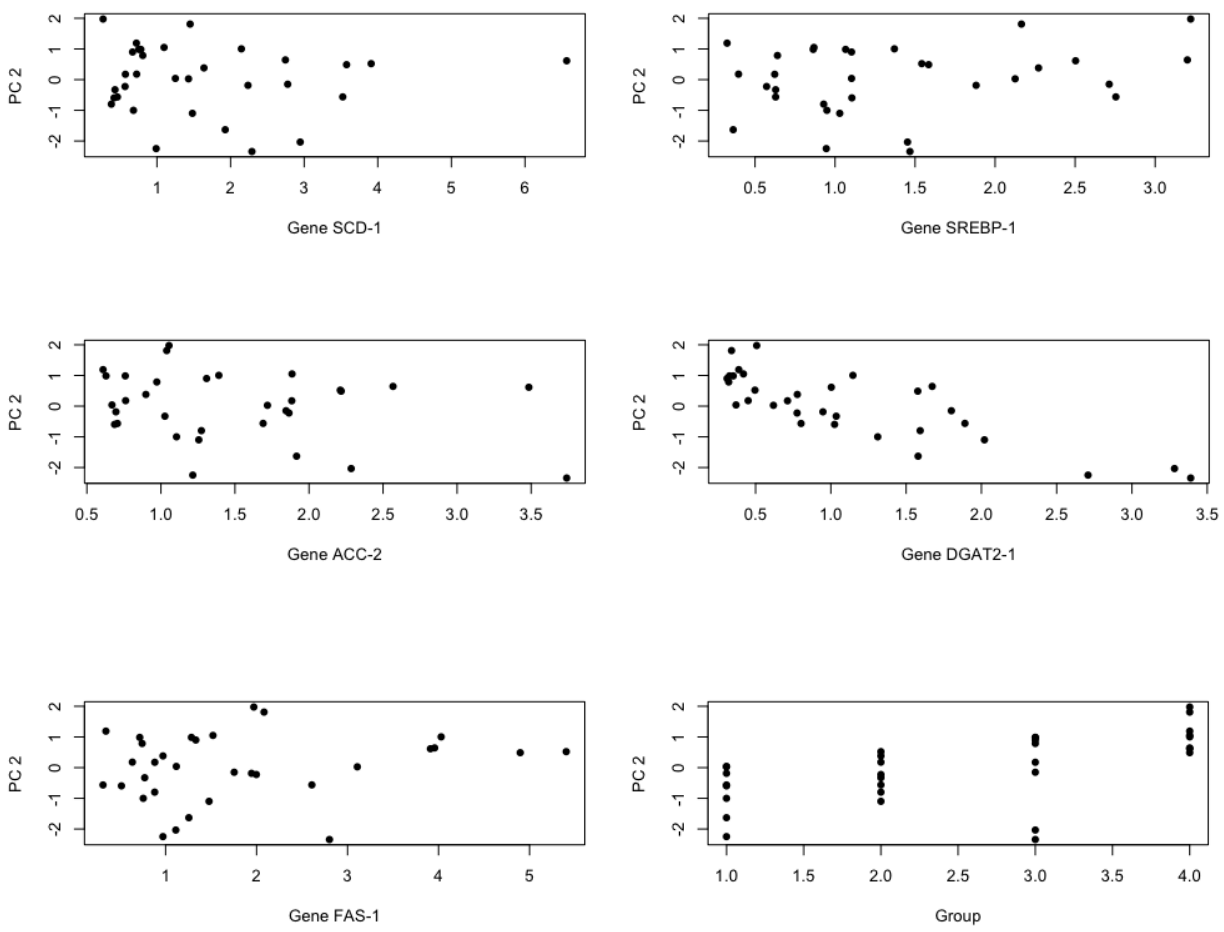

Figure 5: *PC2 versus all variables individually.*

For PC2, DGAT2-1 is the main driving factor with a negative correlation. All other genes have very weak or no correlations. Remembering Figure 2, we noticed that Diet positively affected PC2, which Figure 5 demonstrates.

Discussion

PCA is often used to discover patterns but in our case, we applied it to an experiment that is completed to gain a better understand of what may be going on within the piglets. Not only did a PCA separate our experimental diet groups but also identified how multiple genes contributed to our larger conclusions of dyslipidemia and obesity. To recap, FAS-1, ACC-2, and SCD-1 are all adjacent within the path (see figure 7); SCD-1 is not shown but acts on acetyl-CoA, a product of ACC. SREBP-1 has a master-like control over many gene expressions related to this pathway. Thus, with an upregulation of SREBP-1, it is likely that FAS, ACC, and SCD-1 would be as well. DGAT2-1 is the only gene chosen that is not directly adjacent to another gene of interest. The evidence within Figures 2, 4 and 5 demonstrates that the effect of DGAT2-1 on PC1 and PC2 is different than the rest of the genes. Therefore, we achieved our goal. However, we cannot conclusively say that where DGAT2-1 is within the pathway is the reason for our outcomes. Regarding our hypothesis, yes, it appears that genes nearby within the path have similar gene expressions and, thereby, similar effects on our principal components.

While my analyses went pretty smoothly after acquiring the proper code, things remain unanswered. Scientifically, why is DGAT2-1 so different from the rest? Statistically, we should follow up with a priori test of the diet groups, which allows one tail to identify minor differences that may slip by using a Dunnett's post-test. Overall, we identified correlations that we otherwise missed in our initial analysis. In turn learning there is more than just comparing the means within a section of data.

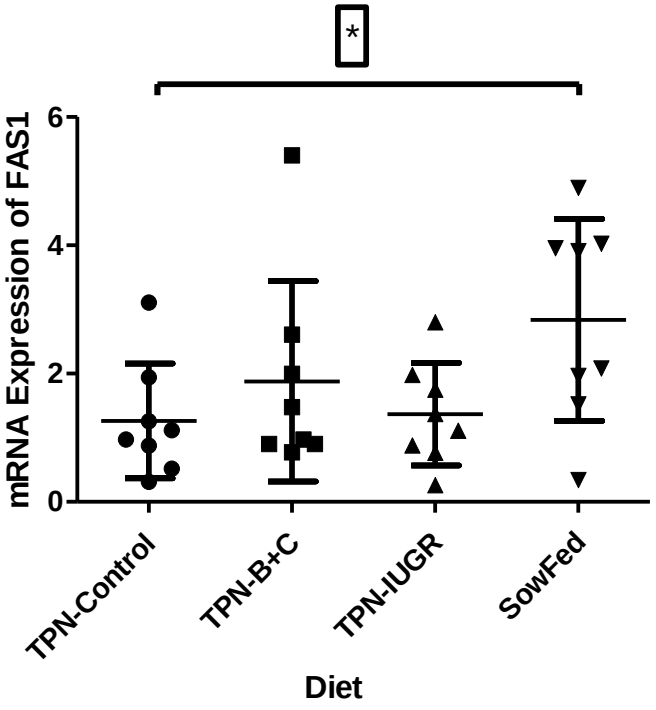

Figure 6: *FAS-1 mRNA expression of four diet groups. One-way anova followed by Dunnett's post test. *statistically significant, p < 0.05.*

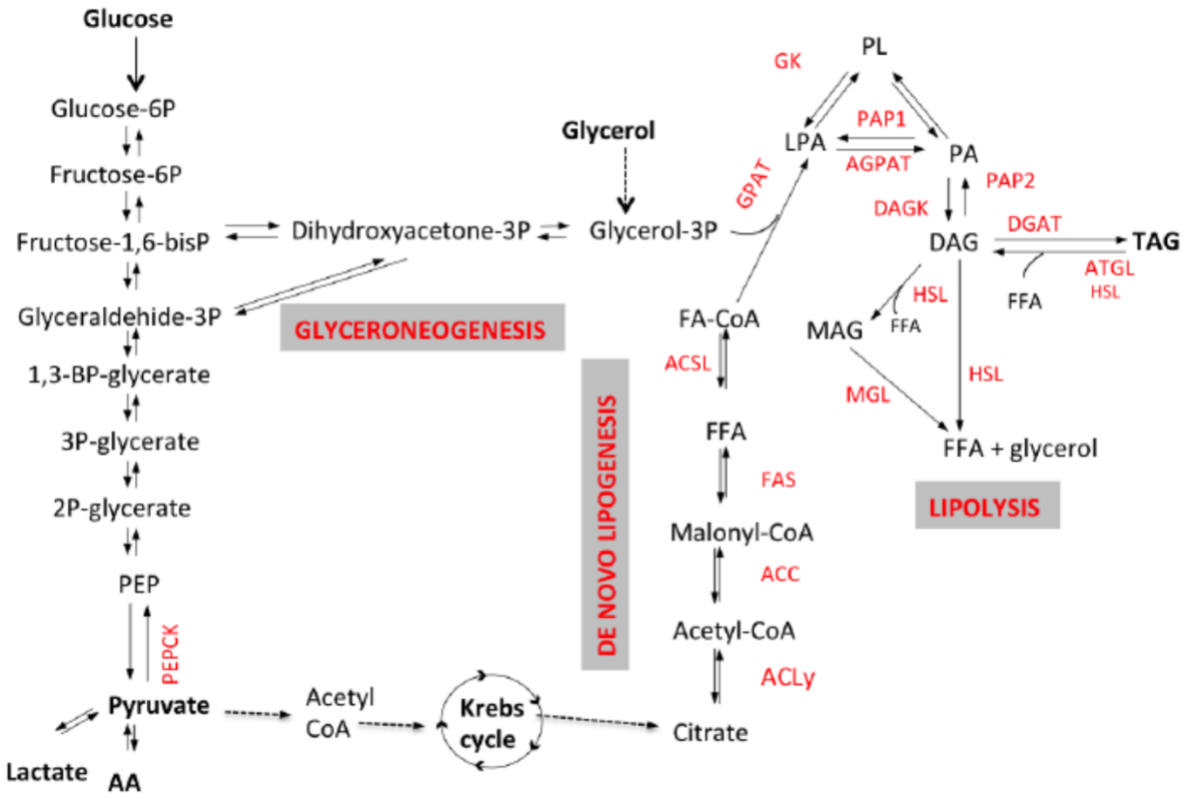

Figure 7: *Lipogenesis pathway for context within discussion. Obtained from [8].*

Code

ChatGpt

```
mRNA <- read.csv("Desktop/MUN/Courses/Stats 7220/Grad
project/mRNA exp2.csv", header = TRUE)
selected_columns <- mRNA[, c("SCD.1", "SREBP.1", "ACC.2",
"DGAT2.1", "FAS.1", "Group")]
pca_result <-prcomp(selected_columns)
summary(pca_result)
eigenvalues <- pca_result$sdev^2
hist(eigenvalues)
pc_scores <- pca_result$x
loading_vectors <- pca_result$rotation
biplot(pca_result, scale = 0)
arrows(0, 0, loading_vectors[, 1], loading_vectors[, 2], angle =
15, length = 0.1, col = "red")
abline(h = 0, lty = 2)
abline(v = 0, lty = 2)
mRNA$Group <- as.factor(mRNA$Group)
plot(pc_scores[, 1], pc_scores[, 2], col = mRNA$Group, pch = 16)
abline(h = 0, lty = 2)
abline(v = 0, lty = 2)
text(pc_scores[, 1], pc_scores[, 2], labels = 1:nrow(pc_scores),
pos = 3)
```

```
legend("topright", legend = levels(mRNA$Group), pch = 1:4, col =
1:4, title = "Groups")
> summary(pca_result)
```

Obtained from Molly Morrissey [7]

```
mRNA<- read_excel("Desktop/MUN/Courses/Stats 7220/Grad
project/mRNA exp.xlsx")
model<-prcomp(mRNA,scale=TRUE)
summary(model)
plot(model,main="Principle Components", xlab = "PC 1-6")
biplot(model)
yv<-predict(model)[,1]
yv2<-predict(model)[,2]
par(mfrow=c(2,2))
plot(mRNA$"SCD-1",yv,pch=16,xlab="Gene SCD-1",ylab="PC 1")
plot(mRNA$"SREBP-1",yv,pch=16,xlab="Gene SREBP-1",ylab="PC 1")
plot(mRNA$"ACC-2",yv,pch=16,xlab="Gene ACC-2",ylab="PC 1")
plot(mRNA$"DGAT2-1",yv,pch=16,xlab="Gene DGAT2-1",ylab="PC 1")
plot(mRNA$"FAS-1",yv,pch=16,xlab="Gene FAS-1",ylab="PC 1")
plot(mRNA$"Diet",yv,pch=16,xlab="Gene Diet",ylab="PC 1")
plot(mRNA$"SCD-1",yv2,pch=16,xlab="Gene SCD-1",ylab="PC 2")
plot(mRNA$"SREBP-1",yv2,pch=16,xlab="Gene SREBP-1",ylab="PC 2")
plot(mRNA$"ACC-2",yv2,pch=16,xlab="Gene ACC-2",ylab="PC 2")
plot(mRNA$"DGAT2-1",yv2,pch=16,xlab="Gene DGAT2-1",ylab="PC 2")
plot(mRNA$"FAS-1",yv2,pch=16,xlab="Gene FAS-1",ylab="PC 2")
plot(mRNA$"Diet",yv2,pch=16,xlab="Diet",ylab="PC 2")
> summary(model)
```

References

[1] "The pig as a research model," RQR. Accessed: Sep. 29, 2021. [Online]. Available: https://www.rqr.umontreal.ca/en/bulletin/the-pig-as-a-research-model/

[2] "Total Parenteral Nutrition (TPN) - Nutritional Disorders," Merck Manuals Professional Edition. Accessed: Oct. 21, 2021. [Online]. Available: https://www.merckmanuals.com/en-ca/professional/nutritional-disorders/nutritional-support/total-parenteral-nutrition-tpn

[3] H. Niinikoski *et al.*, "Onset of Small Intestinal Atrophy Is Associated with Reduced Intestinal Blood Flow in TPN-Fed Neonatal Piglets," *The Journal of Nutrition*, vol. 134, no. 6, pp. 1467–1474, Jun. 2004, doi: 10.1093/jn/134.6.1467.

[4] F. Gao *et al.*, "Marked methylation changes in intestinal genes during the perinatal period of preterm neonates," *BMC Genomics*, vol. 15, no. 1, p. 716, Aug. 2014, doi: 10.1186/1471-2164-15-716.

[5] L. E. McBreairty and R. F. Bertolo, "The dynamics of methionine supply and demand during early development," *Applied Physiology, Nutrition & Metabolism*, vol. 41, no. 6, pp. 581–587, Jun. 2016.

[6] S. D. Bray, "Creatine synthesis and methionine partitioning during TPN in Yucatan miniature piglets," masters, Memorial University of Newfoundland, 2017. Accessed: Oct. 22, 2021. [Online]. Available: https://research.library.mun.ca/13133/

[7] M. M. Morrissey, "Diatom population structure in 58 lakes in the Hudson Bay Lowlands: An Exploratory".

[8] C. Saponaro, M. Gaggini, F. Carli, and A. Gastaldelli, "The Subtle Balance between Lipolysis and Lipogenesis: A Critical Point in Metabolic Homeostasis," *Nutrients*, vol. 7, pp. 9453–9474, Nov. 2015, doi: 10.3390/nu7115475.