5-2016

RNA Seq Analysis of Non-Alcoholic Fatty Liver Disease (NAFLD) Induced by Metabolic Syndrome in a Mouse Model

Diego Almanza
University of Massachusetts Boston

RNA Seq Analysis of Non-Alcoholic Fatty Liver Disease (NAFLD) Induced by Metabolic Syndrome in a Mouse Model

Diego Almanza[1,2], Mehrnaz Gharaee-Kermani[1,2], Jose A. Rodriguez-Nieves, Ph.D.[2] Todd Riley, Ph.D.[1,2], Jill A. Macoska, Ph.D. [1,2]

[1]Department of Biology, University of Massachusetts-Boston, [2]Center for Personalized Cancer Therapy, University of Massachusetts-Boston

Abstract

Metabolic syndrome is often defined by the presence of several factors, including accumulation of abdominal fat, high blood pressure and high blood glucose levels, that increases the risk of developing heart disease, diabetes and cancer. The development of metabolic syndrome has been shown to be very closely linked to lack of physical activity and is increasing concurrent with the rise of obesity rates among adults. Development of metabolic syndrome can have detrimental physiological effects throughout the body and can lead to the development of Non-Alcoholic Fatty Liver Disease (NAFLD). Proper liver function is crucial for the health of the body since it is the site of essential processes such as protein production, blood clotting, metabolism of cholesterol, glucose and iron. Metabolic syndrome contributing to NAFLD can lead to liver dysfunction, hepatitis, cirrhosis, and hepatocellular carcinoma. In this study we used liver RNA from our diet-induce obesity mouse model (mice fed with a 60% fat diet) to characterize the transcriptional landscape of NAFLD and compare it to the transcriptional signature of healthy control mice (mice fed with a 10% fat diet). Understanding the transcriptional differences between these two groups can aid in the understanding of NAFLD and consequent liver pathobiologies. More importantly, the transcriptional signatures identified could also help early detection of these diseases through the identification of novel markers. In this study, we observed significant down-

regulation of genes associated with metabolic processes, and significant up-regulation of genes involved in immune responses. Additionally, we also identified up-regulated transcripts in our obese mice reported to be involved in hepatocellular carcinoma and fibrosis. However, the NAFLD livers did not demonstrate histological evidence of carcinoma and/or fibrosis. Therefore, the NAFLD livers already expressed a premalignant signature in the absence of any other indication of cancer. In conclusion, the identification of genes downregulated in metabolic processes and up-regulated in immune responses indicate that our model is indeed exhibiting liver disease. Moreover, the finding of a premalignant signature suggests that NAFLD may begin to progress towards hepatocellular carcinoma much earlier than previously thought.

Introduction

Metabolic Syndrome (MS) has been steadily increasing as a public concern throughout the years. MS includes a lot of risk associated with it, bringing an increased risk for type 2 diabetes, atherosclerotic cardiovascular disease, and chronic kidney disease(*1*,*2*). The increase of risk factors has been directly associate with glucose intolerance or insulin resistance, obesity, hypertension, proinflammatory state, and dyslipidemia (*2-4*). Many other diseases run parallel to the risk factors mentioned above, therefore, with the rise of obesity in the American population, the prevalence of metabolic syndrome will continue to rise, leading to the cause of other diseases. Worst prognosis such as Non-Alcoholic Fatty Liver Disease (NAFLD), cirrhosis, and ultimately hepatocellular carcinoma can then be direct results from the associated risk factors of metabolic syndrome(*3*). With the rise of obesity in the country, the risk for developing many diseases has been facilitated. Currently, NAFLD has become one of the most common liver diseases and has been estimated that 20% - 30% of individuals in the Western population is diagnosed with NAFLD(*1*, *5*). With increasing amount of individuals being diagnosed with NAFLD, it has become

clear that this disease is not as benign as previously thought, and is now recognized as a major cause for liver-related morbidity and mortality(*1*). NAFLD is characterized with a wide histological spectrum, which includes simple steatosis and non-alcoholic steatohepatitis (NASH). Quantification of the hepatocytes containing fat droplets in the liver can give an indication into the severity of the steatosis, and the recommended lower threshold has been set to 5% of the liver mass being composed of hepatocytes containing fat(*6*). A histological result of above that threshold can result in a diagnosis of steatosis, which can potentially lead to the development of NAFLD. In addition, NAFLD has been directly associated with the same feature as metabolic syndrome, and has be thought to be the hepatic manifestation of this syndrome(*1*). Therefore, it has become a severe problem that needs more focused attention, and a facilitated way of being able to diagnose and perform proper interventions in patients.

It has become clear that the risk factor between metabolic syndrome and NAFLD run parallel to one another, influencing each other to becoming worst prognosis. Obesity is one of these risk factors, and a contributor to many other diseases. Therefore, in our mice model, metabolic syndrome and ultimately NAFLD was induced via different diets, high fat diet (HFD) with 60% of their calories from fat, and low fat diet (LFD) with 10% of their calories from fat. Previous work from our group, showing the effects of obesity in our mice model allowed us to observe the many risk factors associated with metabolic syndrome. This previous study showed that mice after months in the different diets are able to develop insulin resistance, they become obese, and develop other health concerns that significantly reduce the quality of life and health of the mice(*7*). In addition to the already known factors, laboratory testing is used as the main diagnosis tool for determining NAFLD. On the extreme end, liver biopsies are also performed to observe for histological changes with the thresholds mentioned above(*6*). Although there is

diagnostics in place for NAFLD, there are many variables that can affect laboratory testing and its accuracy. Additionally, liver biopsies are also a very invasive method of performing diagnosis. For this reason, transcriptional regulation was of main focus in this study. We aimed to observed the overall regulation of different processes and pathways, and how they differ between the different samples under different diets.

As mentioned before, our group has recently shown that these mice do in fact have the characteristics of metabolic syndrome, and recent histological work has characterized our obese mice with steatosis(*7*). This led us to the conclusion that our mice, aside from having developed metabolic syndrome, they were also diagnoses with NAFLD. Therefore, our study focused in observing the transcriptional differences between our obese mice in a high fat diet (HFD), and lean mice in low fat diet (LFD). We show that there are significant transcript differences between the two phenotypes, up-regulating inflammation-related processes and down-regulating metabolism-related processes in HFD-fed mice compared to LFD-fed mice. Additionally, we also show that although our senescence-accelerated mouse prone (SAMP)6 mice do not show any histological characteristics of cancer/tumorigenesis or fibrosis, they are able to present a pre-malignant and pre-fibrotic signature consistent with early states of tumorigenesis and hepatocellular carcinoma.

Materials and Methods:

Mice Diet:

SAMP6 strain mouse were used in this study, with colonies established from eight females and four males purchased at 6 weeks of age (Harlan Laboratories, Indianapolis, IN) as has been done in previous study from our group(*7*). At 6–8 weeks of age, 25 mice each SAMP6 and AKR/J were fed the following diets: high fat diet (HFD) containing 60% calories from fat, 20% protein, and 20% carbohydrates (58Y1, Test Diet, Richmond, IN); low fat diet (LFD) containing 10.2%

calories from fat, 18.3% protein, and 71.5% carbohydrates (58Y2, Test Diet, Richmond, IN), or regular diet (RD) grain-based mouse chow containing 13.50% calories from fat, 28.50% protein, and 58% carbohydrates (#5001, Lab Diet, St. Louis, MO). The mice were fed daily with fresh high, low fat or regular chow (5 g/day) for 6 months(*7*).

Liver Tissue Harvest:

Mice were sacrificed and the tissue snap frozen in liquid nitrogen. Tissue was immersed in liquid nitrogen and kept there for the remaining of the procedure. Upon the completion of the harvest procedure, the tissues are stored in sterile falcon tubes and later in dry ice. Tissues are then frozen at -70 ^{o}C until the homogenization procedure was performed.

Homogenization Procedure:

Homogenization of the tissue was done using an electric homogenizer. A transfer of at least 1mL TRIZOL reagent (Invitrogen, Carlsbad, CA) per 100mg of tissue to be homogenized was done into a falcon tube and homogenized the tissue until the tissue was completely dissolved in solution.

RNA extraction:

Mixture was vortexed thoroughly. Once homogenized, aliquots of the solution were transferred to Eppendorf tubes and left in TRIZOL at room temp for five minutes. Phase Separation was then performed using 200 µl chloroform per 1mL TRIZOL (originally used), vortexed for 15 seconds, and left at room temp for 5 minutes. Samples were then centrifuged at 12,000g for 15 minutes at 2-8 ^{o}C. RNA Precipitation was done following centrifugation, there were three phases visible within the tube. The aqueous phase (top) was transferred to a fresh RNAse free tube, being careful not to contaminate the solution with the other phases. Added 500 µl isopropanol per 1mL TRIZOL (originally used) to the new tube and incubated at room temp for 10 minutes. Samples

were then centrifuged at 12,000g for 10 minutes at 2-8 ^{o}C. RNA wash and resuspension was done following centrifugation, and the supernatant removed. RNA pellet was washed with 70% cold ethanol in DEPC water / 1ml TRIZOL (originally used) and vortexed. Samples were then centrifuged at 7,500g for 5 minutes at 2-8 ^{o}C. Supernatant was removed. Remaining of ethanol was allowed to air dry for 5 minutes. Pellet was then resolved in 50 µL of DEPC water.

Library Preparation:

Library preparation of the two treated, HFD (n = 3) and LFD (n = 3), SAMP6 mice were done in biological triplicates. Samples were prepared by the Genomics Core in the Center for Personalized Cancer Therapy at UMass Boston. Total RNA samples were DNAse treated and 300ng was used as input to create single indexed (6 base pairs) RNA-Seq libraries using the TruSeq RNA preparation kit according to the manufacturer's protocol(*8*). Samples were pooled together and sequenced on a two-lane flow cell using on-board cluster generation with the Illumina HiSeq^{TM} 2500 instrument.

HiSeq^{TM} 2500:

RNA-Seq libraries were sequenced on the HiSeq^{TM} 2500 loaded with software version 3. A 51 cycle paired-end rapid run of the single indexed RNA-Seq libraries was performed. Raw data of the Bcl base call files were then demultiplexed upon completion.

NanoString:

NanoString nCounter^{TM} Mouse Inflammation v2 panel consisting of 248 inflammation-related mouse genes and 6 internal reference genes was performed in the same total RNA isolated from the liver tissues. Using 100ng input of total RNA was used in preparation for the hybridization reaction following NanoString's nCounter^{TM} Gene Expression Assay Manual(*9*). Post hybridization reaction, samples were then placed in the nCounter Prep Station v4.1.0.1 and

nCounter Digital Analyzer v3.1.0.1 for finalization of the protocol and data collection. Data was analyzed using the nSolver Analysis Software v2.5.

Bioinformatics Analysis:

- *Samples*

  3 high fat and 3 low fat mice were used for the the purposes of this experiment. Sample names from the sequencing were as follows:

  - High Fat mice: NK001_1_CGATGT, NK002_2_TGACCA, and NK003_3_ACAGTG
  - Low Fat mice: NK004_4_GCCAAT, NK005_5_CAGATC, and NK006_6_CCTGTA

- *Bcl basecall files to FASTQ Conversion*

  Bcl basecall files were converted to FASTQ format using Illumina's bcl2fastq conversion tool v1.8.4. Stringent conversion parameters were set by not allowing a mismatch in the index during the demultiplexing process. Additionally, FASTQ files were created overwriting the default parameter to split the files after 4 million reads, generating a single FASTQ file for each sample replicate.

- *Alignment to reference genome*

  Alignment of reads were carried out using Tophat v2.0.14(*10*, *11*) post FASTQ file conversion using Illumina's bcl2fastq tool v1.8.4. Alignment parameters for Tophat consisted of utilizing the default parameters, which were sufficient for downstream analysis of data. The reference genome file used for the alignment was the mm10 Genome Reference Consortium GRCm38.p4(*12, 13*), and the gene transcript file provided by iGenomes (Illumina).

- *Differential Expression Analysis*

  Cufflinks v2.2.1 was used to estimate the relative abundance of transcripts aligned for all the samples(*14*). Cuffdiff v2.2.1 was used to perform the differential gene expression analysis post quantification of the transcripts for the samples(*15*). Analysis was performed against the low fat mice to elucidate up-regulated and down-regulated genes in high fat mice. The statistical cutoff was set to p-value ≤ 0.05.

- *Pathview*

  Pathview v1.10.1 was used in order to perform gene set enrichment analysis to elucidate the top pathways cellular pathways that are overall up-regulated, or down-regulated when compared to LFD(*16*). In addition, selected pathways of interest were also picked for visualization of the overall regulation of the pathways. The statistical cutoff was set to a q-value ≤ 0.05.

- *Network Analysis*

  Network Analysis of the differential expression analysis was done using Cytoscape v3.3.0(*17*). Gene sets of 2-fold or higher genes up-regulated in HFD mice, and 0.1-0.5-fold lower down-regulated genes in the HFD mice with a p-value cutoff of 0.05 was used as the gene set entry for the network analysis of the mouse interactome. Interactome used for this analysis was provided by the Cytoscape program, loaded with the mouse interactome from BioGRID database release 3.4.129.

- *Gene Ontology*

  Gene ontology analysis of the gene sets, both 2-fold or higher genes up-regulated in HFD mice, and 0.1-0.5-fold lower down-regulated genes in the HFD mice were performed using ClueGO and GOstats. ClueGO v2.2.3 was used for the analysis of the two

sets of genes, both up-regulated and down-regulate genes in the HFD mice(*18*). GOstats v2.36.0 was used to observe for overexpression of the biological processes of up-regulated and down-regulated genes in the HFD mice compared to LFD mice(*19*). Statistical cutoff for both computational programs was set to a p-value of ≤ 0.05.

- *nSolver Analysis*

  Raw nCounter data was analyzed using the nSolver Analysis Software v2.5.34. Normalization of the data was carried out in two different steps. First using the geometric mean of the positive controls probes provided by nanoString, the data was normalized to this prior to normalization using the housekeeping probes. Normalized data, and genes with counts higher than 10 were then used for the rest of the analysis. All statistical significance was set to a p-value of ≤ 0.05.

Results

High Fat Induces Metabolic Syndrome Leading to NAFLD

It has been widely shown in the literature that excess of white adipose tissue (WAT) can lead to several health concerns, including inflammation, metabolic syndrome, type 2 diabetes, and NAFLD(*20*). Excess WAT or fat was observed in our SAMP6 mice model as can be seen in the lateral abdominal area of the mice in Figure 1A. The mouse on the left, low fat diet (LFD) mice, shows a relatively normal liver size compared to the mice on the right, high fat diet (HFD), where the liver is significantly bigger in size. This can be a direct effect of the inflammation aspect that that the excess WAT is contributing to. In figure 1B, liver histology images of the mice liver, both HFD and LFD can be observed. Images of the liver histology for HFD-fed mice show various white spotting of fat in the liver. A normal liver's mass consists of about 5-10% of fat, on the other hand a higher fat percentage is considered steatosis as it was identified by the histologist. Lastly,

previous work by our research group was able to identify that the SAMP6 strain mice, under the same dietary conditions, are able to various health concerns(*7*). These mice developed type II diabetes, as previously shown by our research group, a condition that has been tied to having increased risks of developing NAFLD. Taken all together, the SAMP6 mice model used in this study has been able show the diseases associated with metabolic syndrome, and being in different diets, develops to NAFLD.

HiSeq Results Validation Through nanoString Analysis

Two different techniques were used to look into the differential expression in genes. A total RNA HiSeq run was performed to observe the gene expression differences between the two different samples. In addition, a nanoString mRNA inflammation panel was also performed as a validation step for the HiSeq data. In figure 2A-B, we are able to observe a correlation between transcripts both up and down-regulated in our sample dataset. This was very encouraging to observe due to that nanoString mRNA panel was very specific to mRNA inflammation transcripts. Taken all together, the results obtained allowed for the downstream analysis using the HiSeq dataset, which has a wider variety of tools available for carrying out differential analysis. Therefore, data shown in sections below were all carried out using the HiSeq dataset.

Transcriptome Regulation of Inflammatory and Metabolic Response

Differential analysis of these two different conditions was assessed at the gene set level, in order to observe what processes or gene ontology the transcripts are regulating. Two different gene sets were taken for this step of the analysis, we observed the regulation of 2-fold up-regulated and 0.1-0.5-fold down-regulated in HFD-fed mice. Gene ontology was carried out using two different techniques as outlined in the methods section above. Using GOstats, we were able to generate figures 3A and 3C, which shows counts of genes associated with specific biological processes in

the cell. In addition, ClueGO was used to look at gene ontology of biological processes in both gene sets as shown in figure 3B and 3D. In both figures 3A and 3B, the gene counts and gene ontology can be observed for the 2-fold or higher up-regulated genes in the HFD-fed mice compared to LFD-mice. We can observe a strong regulation of genes that are directly associated with response to stimulus and immune response. This up-regulation reflects the metabolic syndrome-induced inflammation in the HFD-fed mice. On the other hand, for gene set of 0.1 to 0.5-fold down-regulated transcripts in HFD-fed mice, the results are shown in figure 3C and 3D. In this figures, we observed that metabolic processes are overall down-regulated in HFD-fed mice when compared to LFD-fed mice. This down-regulation observed reflects the fatty liver organ dysfunction that could have been a direct effect of the HFD that mice underwent during the six months. Taken all together, the up-regulation of immune-related transcripts and the down-regulation of metabolism-related transcripts shows the stress caused by metabolic syndrome-induced inflammation and this paired with the decrease of metabolism function has a significant effect in the health and quality of life of the mice.

Top Regulated Pathways in NAFLD Mice

Additionally, the complete HiSeq dataset of the differential expressed transcripts containing both up-regulated and down-regulated transcripts in HFD mice also underwent gene-set enrichment analysis using the Pathview tool as outlined in the methods above. In figure 4, we have tables of overall up-regulated and down-regulated pathways in HFD-fed mice compared to LFD-fed mice. Figure 4A shows the top significant (p-value ≤ 0.05) overall up-regulated pathways given the transcripts used as input. It can be observed that the “Complement and Coagulation Cascades” came as the top significant up-regulated pathway, which is able to correlate with recent studies. Recent studies by Rensen S. S. *et al*. has shown the implications of complement transcripts

in NAFLD, which is able to support our findings(*21*). In addition, the Complement Cascade has also been shown to be up-regulated in response to inflammation, which reflects the metabolic syndrome-induced inflammation. In addition, many of the pathways also shown in figure 4A has been implicated in inflammation response, once again being able to support the data found during the gene ontology analysis and inflammation of the liver as it was observed in figure 1B.

Gene set enrichment analysis was also performed to look at the overall down-regulated pathways using the Pathview tool. Figure 4B, shows the significantly (p-value ≤ 0.05) down-regulated pathways in HFD-fed mice compared to LFD-fed mice. Once again, just like the gene ontology analysis, we observed a down-regulation of critical pathways implicated in metabolism processes. As previously mentioned, this could be a direct effect of the liver organ dysfunction at not being able to properly metabolize nutrients. Taken this data together, it is able to show the severe effect that HFD-fed mice with NAFLD have, which correlates with current understanding of NAFLD in the patients.

Correlation between NAFLD, Hepatocellular Carcinoma Transcripts and pre-fibrotic signatures

It is well understood that NAFLD can eventually lead to a worst prognosis like cirrhosis and hepatocellular carcinoma if not treated in time or left unattended(*22*). Currently different biomarkers have been associated with Hepatocellular Carcinoma, therefore, we aimed to observe if any of these known biomarkers can be found in our samples when comparing HFD-fed mice to LFD-fed mice(*23*). In figure 5, we are able to observe transcripts that have implications in hepatocellular carcinoma that were also found in our mice model. Performing differential expression analysis in the sample, our group was able to observe the up-regulation of transcripts relevant to hepatocellular carcinoma. Transcripts such as growth factors and heat shock proteins

observed in this figure can directly explain the stress and the over-grown liver observed in figure 1A. The growth factors observed up-regulated in HFD-fed mice provide a pre-malignant signature implicated in cancer, with potential for uncontrollable cell growth, a characteristic unique to cancer(*24*). In addition, our group also observed a strong up-regulation of a pre-fibrotic pathway, the "ECM-Receptor Interaction", as can be observed in figure 6. The ECM has the vital role that contains the information, maintenance, and sustainability necessary for normal function of different tissues(*27*). It is widely understood that up-regulation of extracellular matrix (ECM) proteins have been implicated in fibrosis, and can further the thickening and scarring of connective tissue(*25*, *26*). In figure 6, collagen can be observed to be up-regulated, and recent studies have shown how collagen production and deposition contributes to the development of fibrosis(*27*, *28*). Although our mice do not show fibrotic livers, an up-regulation for collagens can indicate the early phases of fibrosis. Taken this data together, it can be concluded that although our mice do not exhibit any cancerous phenotypes, they are able to present a pre-malignant and pre-fibrotic signatures consistent with early stages of tumorigenesis.

Discussion

Understanding the transcriptional response of diseases can shed light into learning more about different diseases. In metabolic syndrome and NAFLD, there are many contributors to the disease, and having a global understanding of the transcriptional response can aid in the diagnosis of these diseases. Previous research from our group showed that these SAMP6 mice are able to develop type 2 diabetes as a consequence from being obese, one of the many conditions connected to metabolic syndrome, which reduced the quality of life and health of the mice(*4*, *7*). Many studies have associated insulin resistance, obesity, hypertension, proinflammatory state, and dyslipidemia with metabolic syndrome(*2-4*). Many of these factors are able to directly affect other organs, and

for the purpose of this study, we aimed to focus on NAFLD and its effect in the transcriptional regulation of the liver. As stated before, inflammation plays a crucial role in NAFLD and the progression of the disease. Many of the genes up-regulated in HFD-fed mice were stimulus and immune-related when compared to LFD-fed mice. It was encouraging to be able to find in our datasets many immune-related responses in the HFD-fed mice, indicating that this response might be a direct result to the metabolic syndrome-induced inflammation of the liver as was previously indicated. Additionally, we also were able to observe the down-regulation of the metabolic processes-related genes, which was able to indicate the fatty liver organ dysfunction as previously noted. On the other hand, our group was also able to observe what pathways these genes were mainly regulating. In HFD-fed mice, the significant up-regulated "Complement and Coagulation Cascades" pathway was the top up-regulated pathway in our study. Previous studies have implicated the complement-related genes as a response from NAFLD, as described by S. S. Rensen *et al.*(*21*). This finding and the fatty liver histology of the HFD-fed mice containing steatosis allowed us to conclude that our mice were indeed suffering from NAFLD. The immune-response, repressed metabolic processes, and the up-regulation of the complement pathway allows us to observe this unique transcriptional response to NAFLD, which can aid in further understanding this disease. Lastly, the pre-malignant and pre-fibrotic up-regulation of transcripts in the HFD-fed mice allowed us to observe something very unique. Although our mice did not show any cancerous or fibrotic phenotype, there were able to show these transcripts, such as growth factors, heat shock proteins, and collagen which have been implicated in early stages of tumorigenesis and hepatocellular carcinoma(*23*, *27*, *28*). Taken this data together, our group was able to find a transcriptional response and signature unique to NAFLD and very early stages of tumorigenesis. This study was able to provide transcriptional information of the NAFLD disease, and increase

our understanding of the cellular response in mice model, which can be potentially translated to the same response in the human population.

Acknowledgments

Thank you to the members of the CPCT lab for assistance and support. Funding for this research was provided in part by the Sanofi Genzyme grant to the College of Science and Mathematics and NIH/HIDDK U54 DK104310 (J.A.M.). This research was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R25GM076321.

References

1. P. Paschos, K. Paletas, Non alcoholic fatty liver disease and metabolic syndrome. *Hippokratia*. **13**, 9–19 (2009).

2. J. Kaur, A comprehensive review on metabolic syndrome. *Cardiol Res Pract*. **2014**, 943162 (2014).

3. G. Marchesini *et al.*, Nonalcoholic fatty liver disease: a feature of the metabolic syndrome. *Diabetes*. **50**, 1844–1850 (2001).

4. R. H. Eckel, S. M. Grundy, P. Z. Zimmet, The metabolic syndrome. *Lancet*. **365**, 1415–1428 (2005).

5. G. Bedogni *et al.*, Prevalence of and risk factors for nonalcoholic fatty liver disease: the Dionysos nutrition and liver study. *Hepatology*. **42**, 44–52 (2005).

6. S. G. Hubscher, Histological assessment of non-alcoholic fatty liver disease. *Histopathology*. **49**, 450–465 (2006).

7. M. Gharaee-Kermani *et al.*, Obesity-induced diabetes and lower urinary tract fibrosis promote urinary voiding dysfunction in a mouse model. *Prostate*. **73**, 1123–1133 (2013).

8. Illumina, TruSeq Stranded Total RNA Sample Preparation Guide (2012), pp. 1–162.

9. nanoString Technologies, nCounter^{TM} Gene Expression Assay Manual (2008), pp. 1–12.

10. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. **25**, 1105–1111 (2009).

11. D. Kim *et al.*, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36–13 (2013).

12. D. M. Church *et al.*, Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biol*. **7**, e1000112–16 (2009).

13. Mouse Genome Sequencing Consortium *et al.*, Initial sequencing and comparative analysis of the mouse genome. *Nature*. **420**, 520–562 (2002).

14. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*. **28**, 511–515 (2010).

15. C. Trapnell *et al.*, Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology*. **31**, 46–53 (2013).

16. W. Luo, C. Brouwer, Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*. **29**, 1830–1831 (2013).

17. P. Shannon *et al.*, Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*. **13**, 2498–2504 (2003).

18. G. Bindea *et al.*, ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. **25**, 1091–1093 (2009).

19. S. Falcon, R. Gentleman, Using GOstats to test gene lists for GO term association. *Bioinformatics*. **23**, 257–258 (2007).

20. A. Birerdinc, M. Jarrar, T. Stotish, M. Randhawa, A. Baranova, Manipulating molecular switches in brown adipocytes and their precursors: a therapeutic potential. *Prog. Lipid Res.* **52**, 51–61 (2013).

21. S. S. Rensen *et al.*, Activation of the complement system in human nonalcoholic fatty liver disease. *Hepatology*. **50**, 1809–1817 (2009).

22. J. Yu, S. Marsh, J. Hu, W. Feng, C. Wu, The Pathogenesis of Nonalcoholic Fatty Liver Disease: Interplay between Diet, Gut Microbiota, and Genetic Background. *Gastroenterology Research and Practice*. **2016**, 1–13 (2016).

23. T. Behne, M. S. Copur, Biomarkers for hepatocellular carcinoma. *Int J Hepatol*. **2012**, 859076 (2012).

24. M. S. Grandhi *et al.*, Hepatocellular carcinoma: From diagnosis to treatment. *Surg Oncol*. **25**, 74–85 (2016).

25. P. Angulo *et al.*, The NAFLD fibrosis score: A noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology*. **45**, 846–854 (2007).

26. M. Ekstedt *et al.*, Long-term follow-up of patients with NAFLD and elevated liver enzymes. *Hepatology*. **44**, 865–873 (2006).

27. M. A. Karsdal *et al.*, Novel insights into the function and dynamics of extracellular matrix in liver fibrosis. *Am J Physiol Gastrointest Liver Physiol*. **308**, G807–30 (2015).

28. D. A. Brenner *et al.*, New aspects of hepatic fibrosis. *Journal of Hepatology*. **32**, 32–38 (2000).

Figures

Figure Legends:

Figure 1: **A.** Two mice showing the effect of diet in the liver, showing a normal liver (left) and a fatty liver (right). In addition, it can be observed the excess amount of white adipose tissue in the high fat diet mouse. **B.** Histological images of both livers from HFD and LFD mice. Steatosis is evident in the high fat diet liver compared to low fat die mice(*7*).

Figure 2: Both the nanoString (dark grey) and HiSeq (light gray) RNA analysis techniques identify the same up-regulated (A) and down-regulated (B) inflammatory-related genes.

Figure 3: **A-B.** Significant up-regulation of genes directly associated with immune response and response to stimulus. This up-regulation of immune response reflects the metabolic syndrome-induced inflammation. **C-D.** Significant down-regulation of genes directly associated with metabolic processes. This down-regulation of metabolic processes reflects fatty liver organ dysfunction.

Figure 4: **A.** Top significant up-regulated pathways in HFD mice when compared to LFD mice. Up-regulated pathways have implications in inflammatory response and supports the data shown in *Figure 3*. **B.** Top significant down-regulated pathways in HFD mice when compared to LFD mice. Down-regulated pathways have implications in metabolic response and supports the data shown in *Figure 3*.

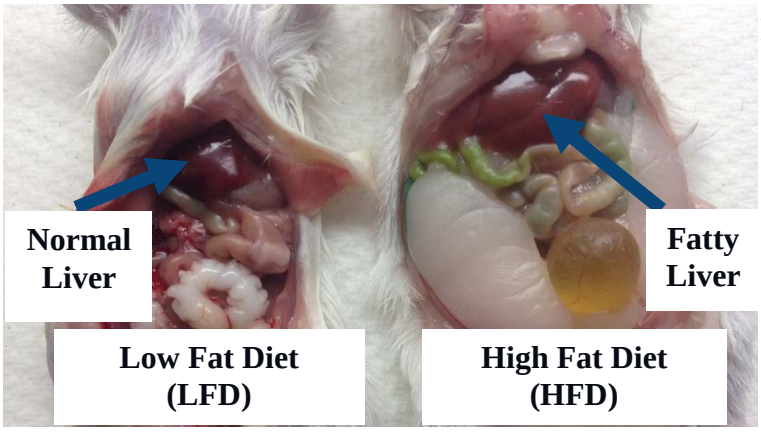
Figure 5: NAFLD Livers Demonstrate a Premalignant Signature. Growth factors and heat shock proteins also found in hepatocellular carcinoma were found up-regulated in HFD mice compared to LFD mice. No tumors were identified in the NAFLD livers. *Therefore, the*

mouse fatty liver demonstrates a pre-malignant signature consistent with early stage of tumorigenesis.

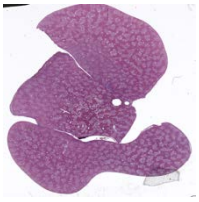
Figure 6: NAFLD Livers Demonstrate a Pre-Fibrotic Signature. We have observed that HFD-fed mice also show an up-regulation for collagen and extra cellular matrix related proteins that have been been implicated in fibrosis. Therefore, our mice model also shows the pre-fibrotic signature that is consistent with early development of cirrhosis and hepatocellular carcinoma.

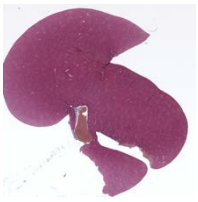
Figure 1:

A.



B.

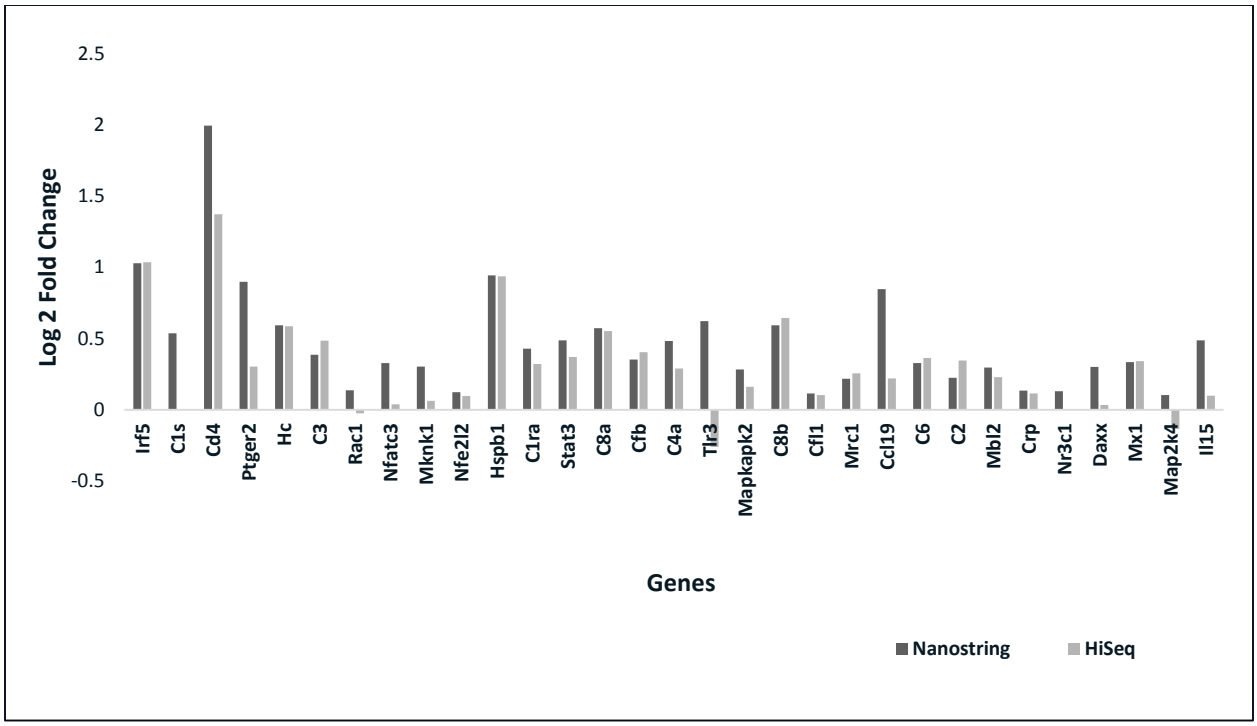




SAMP6 HFD MICE		Histology	RNA	RNA Code
164	321 liver	steatosis	YES	1
165	325 liver	steatosis	YES	2
166	329 liver	steatosis	YES	3
SAMP6 LFD MICE				
992	333 liver	normal	YES	4
993	337 liver	normal	YES	5
994	341 liver	normal	YES	6

Figure 2:

A.


2.5
2
1.5
1
0.5
0
-0.5
Log 2 Fold Change
Irf5
C1s
Cd4
Ptger2
Hc
C3
Rac1
Nfatc3
Mknk1
Nfe2l2
Hspb1
C1ra
Stat3
C8a
Cfb
C4a
Tlr3
Mapkapk2
C8b
Cfi1
Mrc1
Ccl19
C6
C2
Mbl2
Crp
Nr3c1
Daxx
Mx1
Map2k4
Il15
Genes
Nanostring
HiSeq

B.

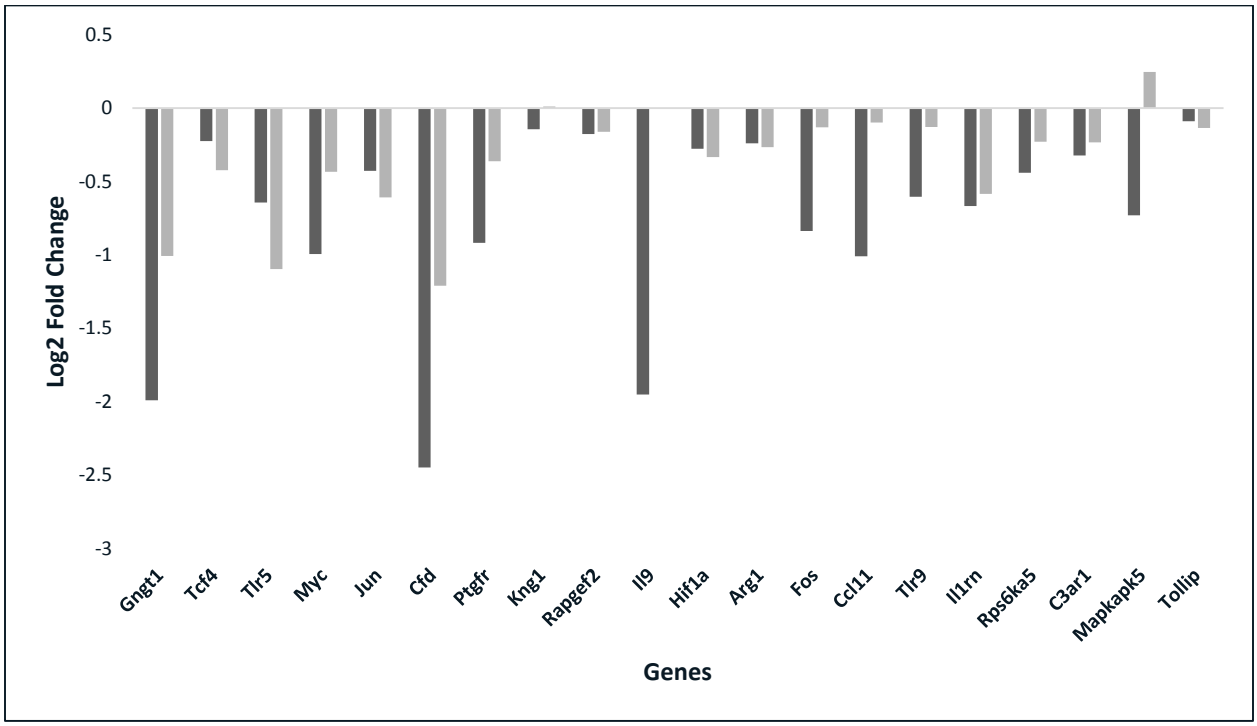
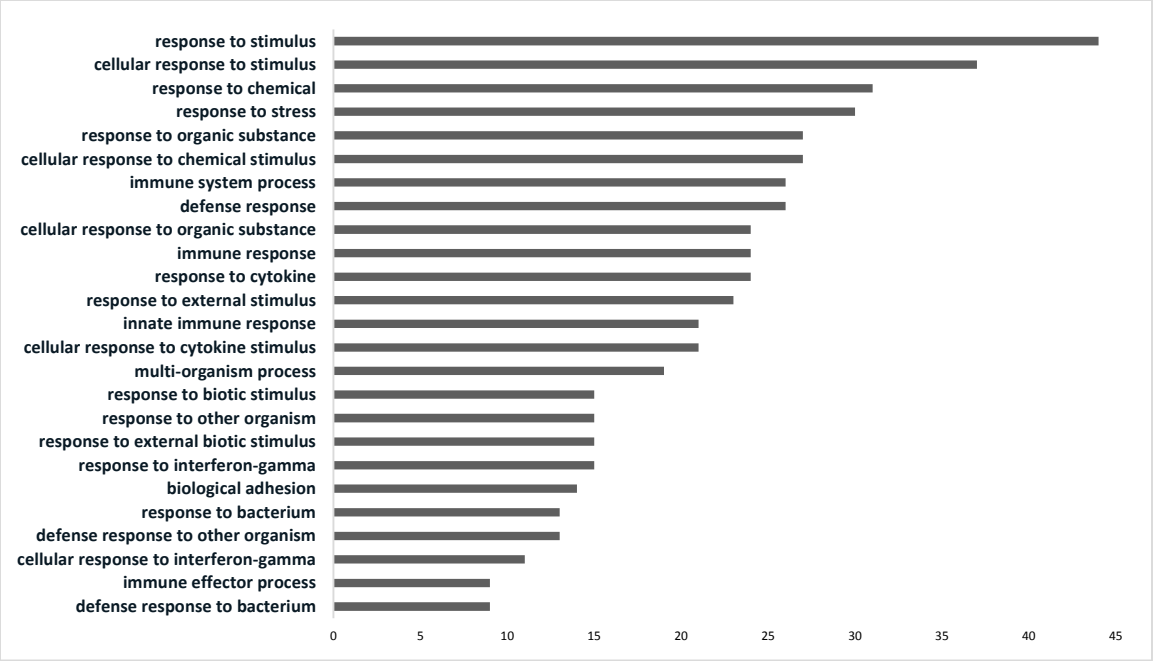
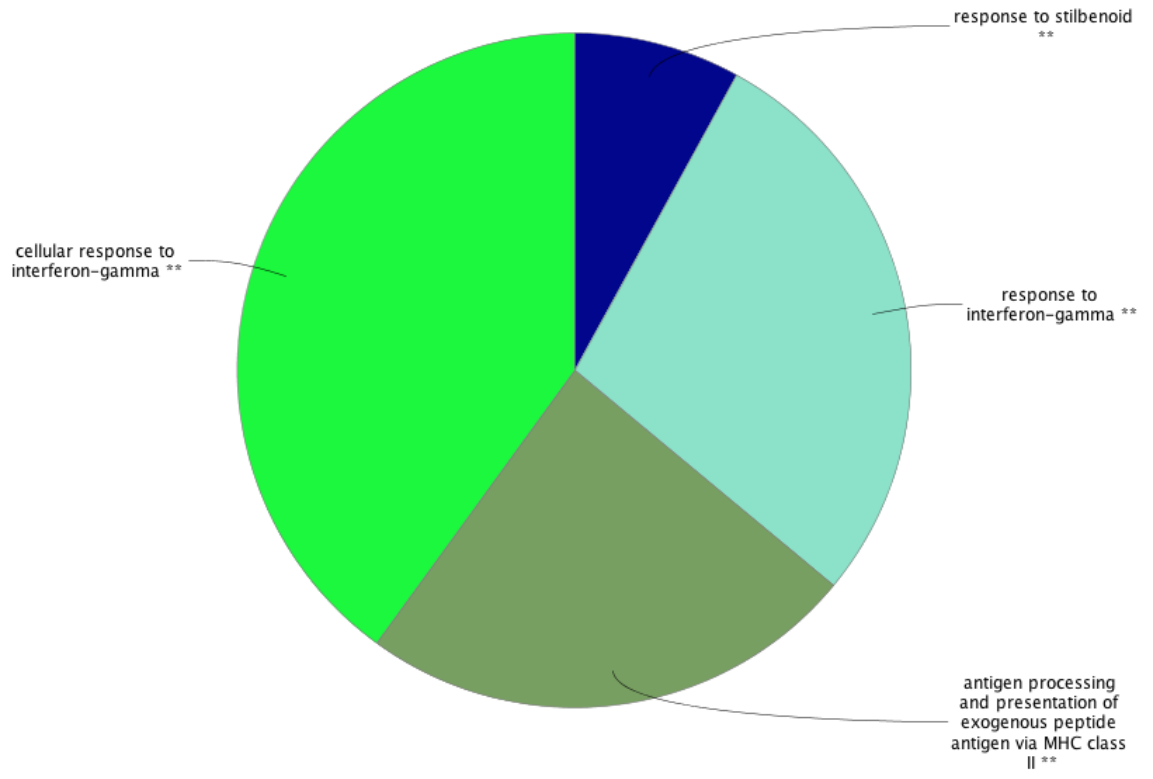

0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
Log2 Fold Change
Gngt1
Tcf4
Tlr5
Myc
Jun
Cfd
Ptgfr
Kng1
Rapgef2
Il9
Hif1a
Arg1
Fos
Ccl11
Tlr9
Il1rn
Rps6ka5
C3ar1
Mapkapk5
Tollip
Genes

Figure 3:

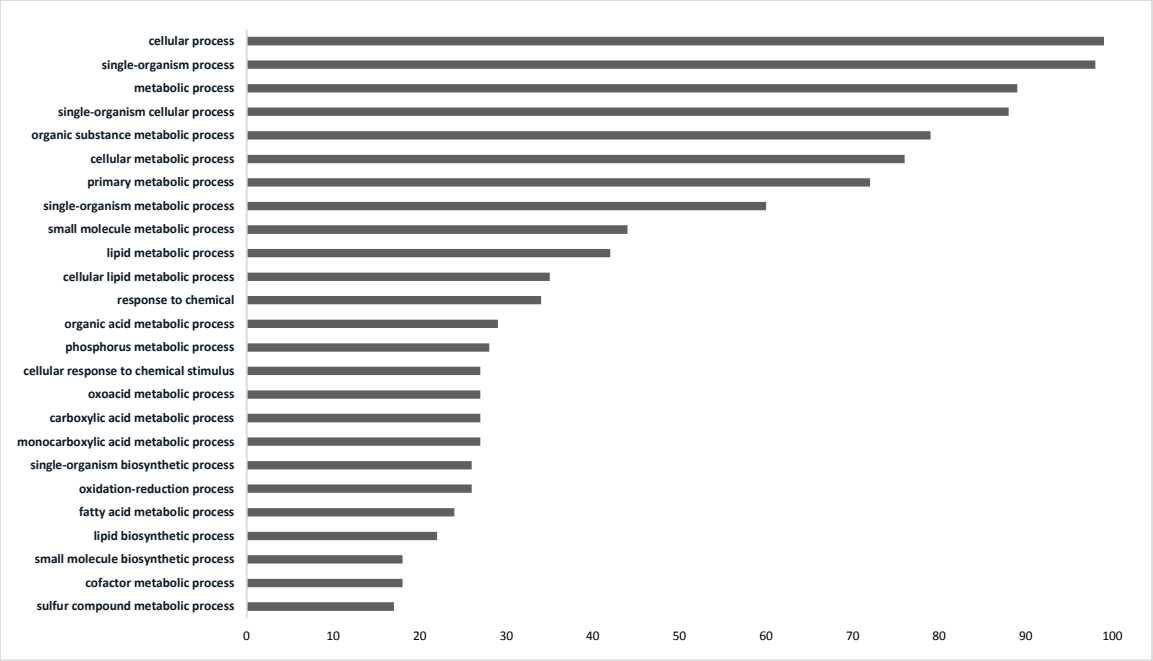
A.


response to stimulus
cellular response to stimulus
response to chemical
response to stress
response to organic substance
cellular response to chemical stimulus
immune system process
defense response
cellular response to organic substance
immune response
response to cytokine
response to external stimulus
innate immune response
cellular response to cytokine stimulus
multi-organism process
response to biotic stimulus
response to other organism
response to external biotic stimulus
response to interferon-gamma
biological adhesion
response to bacterium
defense response to other organism
cellular response to interferon-gamma
immune effector process
defense response to bacterium
0
5
10
15
20
25
30
35
40
45

B.


response to stilbenoid **
response to interferon-gamma **
antigen processing and presentation of exogenous peptide antigen via MHC class II **
cellular response to interferon-gamma **

C.


cellular process
single-organism process
metabolic process
single-organism cellular process
organic substance metabolic process
cellular metabolic process
primary metabolic process
single-organism metabolic process
small molecule metabolic process
lipid metabolic process
cellular lipid metabolic process
response to chemical
organic acid metabolic process
phosphorus metabolic process
cellular response to chemical stimulus
oxoacid metabolic process
carboxylic acid metabolic process
monocarboxylic acid metabolic process
single-organism biosynthetic process
oxidation-reduction process
fatty acid metabolic process
lipid biosynthetic process
small molecule biosynthetic process
cofactor metabolic process
sulfur compound metabolic process
0
10
20
30
40
50
60
70
80
90
100

D.

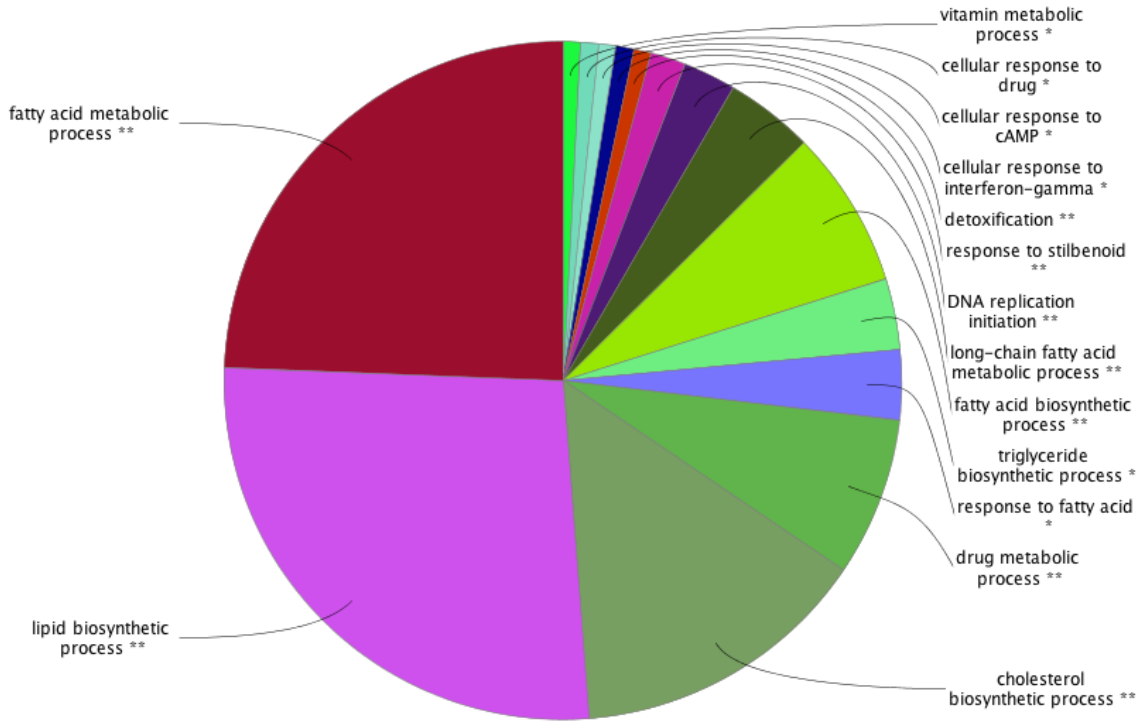

vitamin metabolic process *
cellular response to drug *
cellular response to cAMP *
cellular response to interferon-gamma *
detoxification **
response to stilbenoid **
DNA replication initiation **
long-chain fatty acid metabolic process **
fatty acid biosynthetic process **
triglyceride biosynthetic process *
response to fatty acid *
drug metabolic process **
cholesterol biosynthetic process **
lipid biosynthetic process **
fatty acid metabolic process **

Figure 4:

A.

KEGG Pathview ID	Pathway Name	greater.p.val	greater.q.val
mmu04610	Complement and coagulation cascades	2.21E-08	4.56E-06
mmu04612	Antigen processing and presentation	1.69E-07	1.75E-05
mmu04514	Cell adhesion molecules (CAMs)	8.89E-06	0.000610158
mmu04145	Phagosome	8.37E-05	0.004308377
mmu04623	Cytosolic DNA-sensing pathway	0.000317954	0.013099719
mmu04380	Osteoclast differentiation	0.000984612	0.033805001
mmu04650	Natural killer cell mediated cytotoxicity	0.003006892	0.088488541

B.

KEGG Pathview ID	Pathway Name	p.val	q.val
mmu01040	Biosynthesis of unsaturated fatty acids	6.12E-13	1.26E-10
mmu03320	PPAR signaling pathway	2.69E-12	2.76E-10
mmu01212	Fatty acid metabolism	4.02E-12	2.76E-10
mmu00982	Drug metabolism - cytochrome P450	2.12E-10	1.09E-08
mmu00830	Retinol metabolism	1.81E-08	7.47E-07
mmu00980	Metabolism of xenobiotics by cytochrome P450	2.18E-08	7.49E-07
mmu00071	Fatty acid degradation	3.46E-07	1.02E-05
mmu00062	Fatty acid elongation	6.82E-07	1.63E-05
mmu00100	Steroid biosynthesis	7.13E-07	1.63E-05
mmu04152	AMPK signaling pathway	1.32E-06	2.72E-05
mmu04110	Cell cycle	1.63E-06	3.06E-05
mmu00480	Glutathione metabolism	1.96E-06	3.16E-05
mmu04146	Peroxisome	1.99E-06	3.16E-05
mmu00061	Fatty acid biosynthesis	4.80E-05	0.000706632
mmu00280	Valine, leucine and isoleucine degradation	5.15E-05	0.00070777
mmu00640	Propanoate metabolism	0.000101947	0.001312561
mmu00900	Terpenoid backbone biosynthesis	0.000109874	0.001331414
mmu03030	DNA replication	0.000157732	0.001769226
mmu00040	Pentose and glucuronate interconversions	0.000163181	0.001769226
mmu00053	Ascorbate and aldarate metabolism	0.000194961	0.002008097
mmu00983	Drug metabolism - other enzymes	0.000270501	0.002653488
mmu00020	Citrate cycle (TCA cycle)	0.00033394	0.003126888
mmu04750	Inflammatory mediator regulation of TRP channels	0.00051632	0.00462443
mmu00561	Glycerolipid metabolism	0.000714777	0.00613517
mmu01200	Carbon metabolism	0.000890638	0.007338854
mmu00650	Butanoate metabolism	0.001129557	0.008949564
mmu04972	Pancreatic secretion	0.001323797	0.010100078
mmu04975	Fat digestion and absorption	0.003184941	0.023250792
mmu00620	Pyruvate metabolism	0.00327317	0.023250792
mmu00500	Starch and sucrose metabolism	0.003437088	0.023601341
mmu03440	Homologous recombination	0.003881617	0.025793972
mmu00564	Glycerophospholipid metabolism	0.005546997	0.035708796
mmu00592	alpha-Linolenic acid metabolism	0.007393605	0.045878496
mmu00860	Porphyrin and chlorophyll metabolism	0.007572179	0.045878496
mmu00590	Arachidonic acid metabolism	0.00892375	0.052522642

Figure 5:

gene	locus	sample_1	sample_2	value_1	value_2	FC_HFvsLF	p_value	q_value	significant
Egfr	chr11:16751363-16918352	HF	LF	91.1514	67.444	1.351512366	0.00045	0.00464187	yes
Hgf	chr5:16553494-16623241	HF	LF	12.2895	9.64095	1.274718778	0.00335	0.0244763	yes
Tgfbi	chr13:56609602-56639339	HF	LF	16.4573	12.5851	1.307681306	2.00E-04	0.00230871	yes
Hspa8	chr9:40801272-40805199	HF	LF	545.46	402.378	1.355591011	5.00E-05	0.000666879	yes
Hsp90aa1	chr12:110691035-110696395	HF	LF	69.6731	52.5011	1.327078861	5.00E-05	0.000666879	yes
Hsp90ab1	chr17:45567777-45573261	HF	LF	255.381	199.264	1.281621367	2.00E-04	0.00230871	yes
Hspa5	chr2:34772089-34776529	HF	LF	264.069	204.947	1.288474581	0.00015	0.00180103	yes

Figure 6: