Gene Expression Changes Predict the Severity of NAFLD-like Liver Injury in Male Collaborative Cross Mice

Volodymyr P. Tryndyak[1], Rose A. Willett[1], Suresh K. Nagumalli[1], Mark I. Avigan[2], Frederick A. Beland[1], and Igor P. Pogribny[1]
[1]FDA-National Center for Toxicological Research, Jefferson, AR; [2]FDA-Center for Drug Evaluation and Research, Silver Spring, MD, USA

Abstract

Nonalcoholic fatty liver disease (NAFLD), a prevalent chronic liver disease, is characterized by substantial variations in severity. In this study, we used a genetically diverse Collaborative Cross (CC) mouse population model to analyze the global transcriptome and clarify the molecular mechanisms involved in hepatic fat accumulation that determine the level and severity of NAFLD. Twenty-four strains of male CC mice were maintained on a high fat/high sucrose (HF/HS) diet for 12 weeks and their hepatic gene expression profiles were determined by next-generation RNA-sequencing. The development of the nonalcoholic fatty liver (NAFL) phenotype in CC mice coincided with significant changes in the expression of hepatic genes at the population level, as evidenced by the presence of 724 differentially expressed genes involved in lipid and carbohydrate metabolism, cell morphology, vitamin and mineral metabolism, energy production, and DNA replication, recombination, and repair. Importantly, the expression of 68 of these genes strongly correlated with the extent of hepatic lipid accumulation in the overall population of HF/HS diet-fed male CC mice. Among these genes, the Perilipin 2 (Plin2) and Aldehyde dehydrogenase 3 family member A2 (Aldh3a2) genes, which encode proteins involved in neutral lipid storage (PLIN2) and lipid detoxification (ALDH3A2), exhibited the highest correlation with osmium tetroxide staining, a marker for lipid accumulation. Partial Least Squares (PLS) was performed to identify a set of genes that could be utilized to estimate NAFL severity. PLS modeling using 247 NFL-related genes showed the highest accuracy for predicting NAFL-like liver injury and identifying the individual mouse strains that are highly susceptible to the development of NAFLD induced by a HF/HS diet. A Uniform Manifold Approximation and Projection dimension reduction analysis of the 247 gene set showed a distinct clustering of CC mouse strains characterized by different levels of lipid accumulation in the livers. These findings imply that gene expression profiling combined with a PLS modeling approach may be useful tools to predict NAFLD severity in genetically diverse patient populations and to determine key potential biomarkers of NAFL and potential targets for treatment.

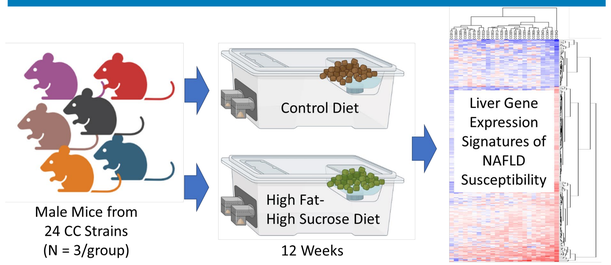
Introduction

Nonalcoholic fatty liver disease (NAFLD) has grown in global frequency to become the most common chronic liver disease, with a prevalence ranging from 25% to 48% in adults and from 8% to 12% in children. NAFLD is a biologically and clinically heterogenous condition encompassing a spectrum of liver-related pathologies ranging from hepatic steatosis without inflammatory changes (nonalcoholic fatty liver, NAFL) to nonalcoholic steatohepatitis (NASH), with inflammation and hepatocyte injury, to fibrosis and cirrhosis. Current evidence indicates the existence of substantial interindividual heterogeneity in susceptibility to NAFLD and its severity. These interindividual differences are influenced by several factors, including a number that have yet to be fully elucidated. The development of new biomarkers that accurately identify NAFLD-prone individuals and NAFLD patients who are at greatest risk for the development of an advanced form of the disease would represent an important advance in the practice of modern medicine.

Molecular drivers of NAFLD are multifactorial and different metabolic and cellular pathways are involved in NAFLD development and progression. Several molecular alterations in NAFL and NAFLD have been implicated by transcriptomic studies as potential biomarkers. These alterations have been associated with some critical disease-related pathway changes tied to NAFLD development and stratification. Furthermore, transcriptomic studies have also been successfully applied for the identification of sex-specific differences in NAFLD.

In this study, we have used a global transcriptomic approach to investigate gene expression profiles associated with the development of NAFLD induced by an obesogenic diet in genetically diverse CC mice and identify the subsets of genes whose levels of expression are associated with the severity of hepatic steatosis.

Materials and Methods



Animals, experimental design, and transcriptomic analysis. Male mice of 24 CC mouse strains were subjected to dietary NAFLD induced by feeding a high-fat and high-sucrose (HF/HS) diet for 12 weeks. RNA libraries for RNA-sequencing were prepared from hepatic RNA using TruSeq Stranded Total RNA library preparation kits with Ribo-Zero Gold for rRNA depletion. Mouse intra-strain differential gene expression analysis was performed using Illumina DRAGEN RNA and Differential Expression software, which uses the DESeq2 algorithm. To minimize inter-strain differences and batch effects, the raw read counts were normalized using ComBat-seq followed by normalization using DESeq2 software.

For DNA methylation genomic DNA from liver samples was digested with the MspI restriction enzyme followed by DNA bisulfite conversion using Zymo Research EZ DNA Methylation-Gold kits. Genomic libraries were prepared using Accel-NGS Methyl-Seq DNA library kits for DNA methylation next-generation sequencing (NGS). R package MethylKit was used for the differential CpG site DNA methylation analysis.

Partial least-squares modeling. Batch-corrected normalized filtered log-transformed gene expression data (a total of 11079 genes) were used for modeling analysis. Supervised predictive models were constructed using Partial Least Squares (PLS) regression, with osmium staining values as the Y variable. The critical variables (genes) for modeling were identified by Variable Importance in Projection (VIP) values, which are the weighted sum of squares of PLS weights and represent the influence of each variable of the data matrix on the model of the response matrix.

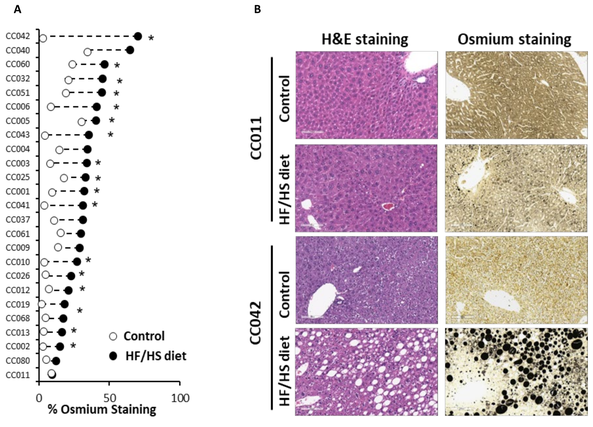


Figure 1. Effect of a high-fat and high-sucrose diet on metabolic alterations and extent of steatosis across the Collaborative Cross mouse strains. (A) Strain-specific changes in osmium staining in HF/HS diet-fed mice and control diet-fed mice. (B) Representative hematoxylin and eosin and osmium staining of liver sections from CC011 and CC042 male mice

Results and Discussion

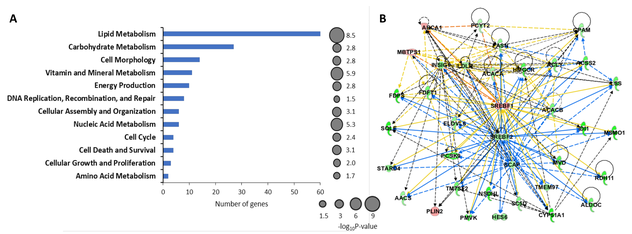


Figure 2. Effect of a HF/HS diet on the gene expression profiles in the livers of male mice from 24 Collaborative Cross (CC) mouse strains. A: Pathway analysis of differentially expressed genes (DEGs) in the livers of male CC mice. B: Visualization of the gene network of the most dysregulated molecular lipid metabolism pathway.

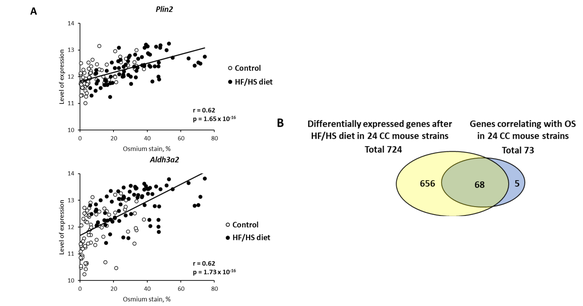


Figure 3. Association of gene expression with NAFL-related alterations in the livers of CC mice fed a HF/HS diet. A: Correlation of Plin2 and Aldha3a2 gene expression with osmium tetroxide staining in the livers of CC mice fed control and HF/HS diets. B: Venn diagram of DEGs in the overall CC mouse population and genes that correlated with osmium tetroxide staining.

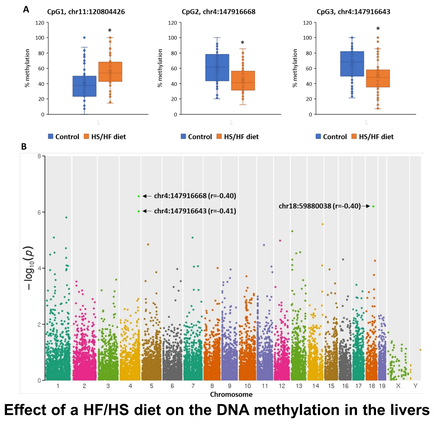


Figure 4. Effect of a HF/HS diet on the DNA methylation in the livers of male mice from 24 CC mouse strains. A: Box plots for percentage of three significantly differentially methylated CpG sites in male mice fed a control diet or HF/HS diet. B: Manhattan plot of correlation data between CpG sites methylation and osmium tetroxide staining in the livers of CC mice fed a control and HF/HS diet.

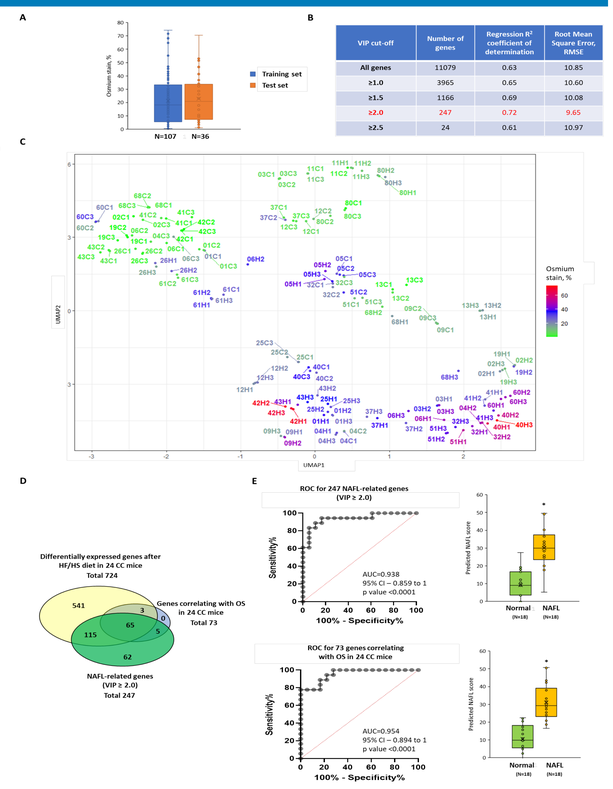


Figure 5. Development of the partial least-squares (PLS) models to predict NAFL-related alterations in the livers of Collaborative Cross mouse strains fed a HF/HS diet. A: Box plot of hepatic osmium tetroxide staining in the "training" (107 samples) and "test" (36 samples) groups of CC mice. B: PLS model evaluation. C: UMAP analysis of 247 NAFL-related genes with VIP ≥ 2.0 in the livers of male mice from 24 CC mouse strains fed a control and HF/HS diet. D: Venn-diagram of 724 DEGs at population level, 73 genes that significantly correlated with osmium staining, and 247 NAFL-related genes with VIP ≥ 2.0 in the livers of 24 CC mouse strains. E: A receiver operating characteristic (ROC) curve analysis of predicted NAFL scores in the "test" group of CC mice.

Conclusion

Extensive alterations in the expression of disease-related genes are a fundamental feature of the pathogenesis of NAFL in a genetically diverse CC mouse population. Importantly, these transcriptomic changes are associated with, and may contribute to, different levels of severity of this fat-accumulating condition in hepatocytes. These findings indicate that analysis of gene expression may be a useful tool in NAFLD detection, especially to identify vulnerable individuals in genetically diverse populations. This concept is further supported by the results of PLS modeling that demonstrate the power of gene expression-based PLS NAFL modeling to identify those individual mouse strains across genetically diverse CC mice that are susceptible to the development of NAFL induced by a HF/HS diet.