8-2012

Population-based risk factors for elevated alanine aminotransferase in a South Texas Mexican-American population

Hui-Qi Qu

Quan Li

Megan L. Grove

Yang Lu

Jen-Jung Pan

See next page for additional authors

Published in final edited form as:
Arch Med Res. 2012 August ; 43(6): 482–488. doi:10.1016/j.arcmed.2012.08.005.

Population-based Risk Factors for Elevated Alanine Aminotransferase in a South Texas Mexican–American Population

Hui-Qi Qu[a], **Quan Li**[b], **Megan L. Grove**[c], **Yang Lu**[a], **Jen-Jung Pan**[d], **Anne R. Rentfro**[e], **Perry E. Bickel**[f,g], **Michael B. Fallon**[d], **Craig L. Hanis**[c], **Eric Boerwinkle**[c], **Joseph B. McCormick**[a], and **Susan P. Fisher-Hoch**[a]

[a]University of Texas School of Public Health, Brownsville, Texas

[b]Endocrine Genetics Lab, The McGill University Health Center (Montreal Children's Hospital), Montréal, Québec, Canada

[c]University of Texas Health Science Center, School of Public Health, Human Genetics Center, Houston, Texas

[d]Division of Gastroenterology, Hepatology and Nutrition, University of Texas Health Science Center at Houston, Houston, Texas

[e]University of Texas at Brownsville and Texas Southmost College, College of Nursing, Brownsville, Texas

[f]Center for Metabolic and Degenerative Diseases, Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, Texas

[g]Division of Endocrinology, Diabetes and Metabolism, Department of Internal Medicine, University of Texas Health Science Center at Houston, Houston, Texas

Abstract

Background and Aims—Elevated alanine aminotransferase (ALT >40 IU/mL) is a marker of liver injury but provides little insight into etiology. We aimed to identify and stratify risk factors associated with elevated ALT in a randomly selected population with a high prevalence of elevated ALT (39%), obesity (49%) and diabetes (30%).

Methods—Two machine learning methods, the support vector machine (SVM) and Bayesian logistic regression (BLR), were used to capture risk factors in a community cohort of 1532 adults from the Cameron County Hispanic Cohort (CCHC). A total of 28 predictor variables were used in the prediction models. The recently identified genetic marker rs738409 on the *PNPLA3* gene was genotyped using the Sequenom iPLEX assay.

Results—The four major risk factors for elevated ALT were fasting plasma insulin level and insulin resistance, increased BMI and total body weight, plasma triglycerides and non-HDL cholesterol, and diastolic hypertension. In spite of the highly significant association of rs738409 in females, the role of rs738409 in the prediction model is minimal, compared to other epidemiological risk factors. Age and drug and alcohol consumption were not independent determinants of elevated ALT in this analysis.

Address reprint requests to: Hui-Qi Qu, Ph.D., Assistant Professor, Division of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas School of Public Health, Brownsville Regional Campus, Brownsville, Texas; Phone: 956 882 7006; FAX: 956 882 5152; huiqi.qu@uth.tmc.edu.

Conclusions—The risk factors most strongly associated with elevated ALT in this population are components of the metabolic syndrome and point to nonalcoholic fatty liver disease (NAFLD). This population-based model identifies the likely cause of liver disease without the requirement of individual pathological diagnosis of liver diseases. Use of such a model can greatly contribute to a population-based approach to prevention of liver disease.

Keywords

Alanine aminotransferase; Liver disease; Machine learning; NAFLD; *PNPLA3* polymorphism; Public health

Introduction

High rates of chronic end-stage liver disease have been documented together with significantly elevated prevalence of diabetes and obesity among Mexican–Americans living at the United States (U.S.)/Mexico border (1–3). Most striking are data from a randomly recruited cohort from this population in which we show a high rate (~39%) of the metabolic syndrome and elevated alanine aminotransferase (ALT) levels, indicative of liver injury (4). We observe a marked gender effect with young males more likely to be obese and to have raised ALT levels (5). These rates are in the absence of evidence for excessive alcohol consumption, but in any event alcoholic and nonalcoholic fatty liver disease (NAFLD) are not exclusive processes and may be additive. These observations raise the important question as to whether this population has high rates of NAFLD and, more importantly, nonalcoholic steatohepatitis (NASH), which leads to end-stage liver disease. Although elevated ALT is known to be indicative of liver injury, it lacks diagnostic specificity for NAFLD in the absence of liver biopsy. Because the risks and the cost of liver biopsy, particularly in a disadvantaged population, are prohibitive on a large scale, we applied machine learning methods to our database in order to identify risk factors for the elevated ALT from extensively documented clinical and biological information. From this we obtain an estimate of the potential burden of NAFLD in our population of Americans of Mexican descent. This knowledge is important in health disparity populations ill equipped to bear additional burdens of preventable liver disease both economically and socially.

The Hispanic population in the city of Brownsville, Cameron County, Texas is one of the poorest in the U.S. (2). Since 2003 we have recruited >2000 healthy participants randomly selected from the community: the Cameron County Hispanic Cohort (CCHC) (2). These individuals consented to extensive sociodemographic, anthropometric and biological analyses. Using weighted data (i.e., data corrected for sampling bias based on census data to account for age, gender, tract/block and household clustering) from this cohort show the prevalence of obesity and diabetes to be high; 7.9% individuals are morbidly obese (BMI ≥40), 48.5% individuals overall are obese (BMI ≥30), and 81.7% are obese or overweight (BMI ≥25). Using the 2010 definition of diabetes recommended by the American Diabetes Association (ADA) (6), 19.2% had pre-diabetes [fasting plasma glucose (FPG) 100–125 mg/dL or A1C 5.7–6.4%] and 30.7% had diabetes (FPG ≥126 mg/dL or glycosylated hemoglobin [HbA1c] ≥6.5%). In addition to obesity and diabetes, we found the prevalence of elevated ALT (≥40 U/L) to be 40.8%. ALT is mainly produced in the liver and released into the bloodstream as the result of liver injury. Chronic liver disease is one of the major causes of death in the adult Hispanic population in the U.S. (2008 statistics of the National Center for Injury Prevention and Control: NCIPC, http://www.cdc.gov/injury/wisqars/fatal.html). Liver disease is ranked as the eighth leading cause of death among Hispanics aged 25 to 34 years, rising to sixth at 35–44 years, and fourth between the ages of 45 and 64 years. Our own data yielded rates of 126/100,000 for end-stage liver disease in a retrospective chart review using ICD-9 codes for end-stage liver disease (1). Rates were considerably higher in males (386/100,000) and overall 8.7% of the 176 cases identified had been diagnosed with hepatocellular carcinoma. There were no biopsy data and only four patients had been referred for liver transplant and none had received one. This study drew charts from a Federally Qualified Clinic in Brownsville serving the same mainly uninsured Mexican–American population from which we drew the CCHC (1). As stated above, the limitation of all these data is that accurate diagnosis of the cause of liver disease depends on the relatively invasive and expensive procedure of liver biopsy. Given these constraints and the concerns raised by our data, we sought to generate more precise data on risk factors using less invasive procedures (venipuncture).

Recently, genome-wide association studies have provided a new tool in the identification of genetic susceptibility of liver injury (7,8). Two single nucleotide polymorphisms (SNP) rs738409 (causing the amino acid substitution Ile148Met) and rs2281135 in the *PNPLA3* locus have been highlighted as being associated with NAFLD (7) and elevated ALT (9), respectively. Our preliminary genetic analysis of the CCHC suggests rs738409 tags the genetic association with elevated ALT better than rs2281135. In addition, our data showed the genetic susceptibility tagged by rs738409 was not biased by population structure of the admixed Mexican–American population. Therefore, genotypes of rs738409 were used as a genetic risk factor in the machine learning process.

Machine learning methods are able to automatically capture risk factors from a large number of variables. The support vector machine (SVM) and Bayesian logistic regression (BLR) are the two most representative machine learning methods for disease risk modeling. SVM is a modern machine learning method that operates by finding an optimal separating hyperplane between affected and unaffected individuals (10). SVM is particularly useful in classifying high-dimensional data and taking into account the interactions among environmental and genetic factors (11,12). Logistic regression is a classical method in disease risk modeling. BLR extends the logic regression to a Bayesian framework by incorporating prior information (13). In this study we aimed to identify the risk factors contributing to elevated ALT in the Brownsville CCHC using these two machine learning methods. We anticipate that this approach will provide a robust measure of the likely disease processes associated with abnormal liver function in this Mexican–American population without invasive liver biopsy. The information will be important for public health policy makers and planners to develop the most efficient prevention and disease management strategies at the population level.

Materials and Methods

Ethics Statement

Written informed consent was obtained from each participant, and the study was approved by the Committee for the Protection of Human Subjects of the University of Texas Health Science Center at Houston (UTHealth).

Subjects

This study investigated 1532 adult individuals on whom we had complete data, recruited prospectively in the Cameron County Hispanic Cohort (CCHC). These individuals were from households randomly selected for recruitment on the basis of 2000 census tract data in the city of Brownsville, Cameron County, Texas (2). The general description of this cohort is in our previous report (2). Among these adult participants in this study, 39% have ALT >40.

Data Management and Confidentiality

Standard protocols for data entry, cleaning, and quality control of the data were applied throughout. Personal identifiers are secured separately with access limited to only those personnel needing to contact participants who had given prior consent to be recontacted. De-identified data are also secured behind the UTHealth firewall with access only to approved collaborators. Data weighting was performed as described (2).

Genotyping

The genotyping of rs738409 was performed using the Sequenom iPLEX assay (Sequenom, Cambridge, MA). The genotyping call rate was 100%. For the purpose of quality control, 93 DNA samples were genotyped in duplicate. The concordance rate of each duplicate is 100%.

Variables for Risk Modeling

The following variables were included in our risk model: gender, age, rs738409 genotype, body mass index (BMI), body height, waist circumference, hip circumference, waist/hip ratio, pulse rate, blood pressure, physical activity, alcohol consumption, smoking, education levels, status of diabetes (diagnosed by the ADA 2010 guidelines) (6), history of hepatitis, medications, fasting plasma glucose (FPG), fasting plasma insulin level, homeostasis model assessment-estimated insulin resistance (HOMA-IR) (14), fasting lipids [serum triglycerides, high-density lipoprotein cholesterol (HDL-c), non-HDL-c, and low-density lipoprotein cholesterol (LDL-c)]. These input variables were linearly scaled to the range [0; +1] and were mapped into a high-dimensional feature space.

Machine Learning Methods

In this study, all classification tasks were performed by support vector machine (SVM) and Bayesian logistic regression (BLR). SVM is a very effective supervised machine learning classifier widely used in pattern recognition or classification. Our soft margin SVM model was implemented with the LIBSVM package (http://www.csie.ntu.edu.tw/~cjlin/libsvm) (15). The radial basis function (RBF) kernel was chosen in this study, which gives the highest accuracy for our test. In our study, the RBF kernel showed better performance than the linear kernel in the SVM model (AUC scores: 0.743 vs. 0.735 in males; 0.690 vs. 0.664 in females). For parameter selection, a grid search heuristic was imposed with 10-fold cross-validation. The weight (or relative importance) of each variable in the SVM model was assessed by the F-score. The F-score measures the discrimination of two groups where the larger F-score suggests the feature (elevated ALT in this paper) and is better discriminated by the variable (16). BLR is the extension of binary logistic regression model. Compared to a standard logistic regression model, the regression coefficients in BLR were estimated with Bayesian prior density (13). Our BLR model was implemented with the Laplace prior part of the Bayesian binary regression (BBR) software (http://code.google.com/p/bbrbmr/). The weight of each variable in the BLR model was assessed by the maximum likelihood β value. Comparisons between the two groups (normal ALT ≤40 vs. elevated ALT >40) were performed using Student t-tests for continuous variables and Pearson χ^2 tests for categorical variables. For the purposes of modeling, we chose this cut-off for ALT so that our results would be comparable with a previous study using data from a different population (17). This study used the receiver operating characteristic (ROC) curve to assess the model performance. The ROC curve plot was generated by calculating true-positive rates and false-positive rates over a relevant range of thresholds. For obtaining aggregate numbers of true vs. false positive rates, the thresholds of each ROC curve underwent stepwise variation from 0–1 in each 0.01 interval. Each threshold was assigned as the probability of an individual having liver injury. The area under the curve (AUC) was used as a measure of performance of the classifiers.

Results

We modeled the risk of liver injury for males and females separately because the rate of elevated ALT is highly stratified according to gender (male = 54.7%, female = 27.4%). In the risk modeling, the performances of the SVM method and the BLR method were assessed based upon area under the receiver operator characteristic curve (AUROC) scores of 0.743 (males) and 0.690 (females) for SVM, 0.693 (males) and 0.670 (females) for BLR (Figure 1). Although SVM and BLR had different performances in our study, SVM results were largely supported by BLR results. The weights of predictor variables in the SVM model represented by the F-scores were validated by removal of a specific predictor variable and then reassessment of AUC scores. For example, consistent with the F scores of the genetic marker rs738409, AUC scores before and after removing rs738409 in the SVM model were 0.743 vs. 0.743 in males and 0.690 vs. 0.669 in females. The risk factors contributing to elevated ALT in males and females are shown in Table 1. Our models identified four groups of risk factors in both genders: (1) abnormal fasting plasma insulin level and insulin resistance (HOMA-IR); (2)(3) serum triglycerides and non-HDL-c; (4) diastolic hypertension; and interestingly (5) genetic susceptibility tagged by the *PNPLA3* SNP rs738409 in females. Alcohol consumption was not identified as a risk factor in this population. These risk factors of increased ALT in this community cohort are consistent with the known risk factors of NAFLD (17,18).

Increased glutamate oxaloacetic transaminase (also known as aspartate aminotransferase, AST) may also manifest liver injury. Elevated ALT with ALT/AST <1 are considered counter to the diagnosis of NAFLD (19). Among the 556 individuals with ALT >40, there are 93 individuals with ALT/AST <1. Our risk modeling in this subset of 93 subjects showed poor AUROC scores suggesting that the risk factors in these participants were largely unrelated to metabolic syndrome, and therefore undetermined using the current set of variables. Nevertheless, our risk model in these 93 individuals did capture increased insulin level and HOMA-IR as the major risk factors in males, with history of hepatitis as the second risk. In females, our modeling highlights the major risk from BMI and waist circumference, whereas non-HDL-c is the second risk. Even in participants with elevated ALT and ALT/AST <1, metabolic syndrome remains a risk factor for liver injury.

Discussion

This study highlights the major but under-appreciated health and economic threats of liver disease in disadvantaged populations with high rates of obesity and diabetes. Obesity induces NAFLD through dysfunctional adipose metabolism mediated by adipokines (20). Our results show that body weight consistently contributes particular risk of liver injury independent of BMI in both males and females. These observations suggest that total amount of body fat is of special importance in contributing to the risk of NAFLD. Furthermore, waist circumference and waist/hip ratio also contribute to liver injury, emphasizing the importance of central fat distribution. Because our observations are highly statistically significant, they underline recent data regarding the roles of central and generalized obesity in NAFLD (21). Although generalized obesity increases the risk of NAFLD, central obesity makes an additional and independent contribution to NAFLD. We confirm that insulin resistance and increased fasting plasma insulin are also important correlates of liver injury, most markedly in females. To date, it is still unclear whether NAFLD is caused by insulin resistance (22) or leads to insulin resistance because of critical disturbances in liver metabolism (23). However, dysfunctional adipose metabolism is a common pathological mechanism shared by insulin resistance and NAFLD (24,25).

Both serum triglycerides and non-HDL cholesterol contributed to the risk of liver injury in this population. The risk effect of non-HDL-c is especially obvious in males manifested by its high rank among all the risk factors (Table 1). This differs from the previously reported correlation between hypertriglyceridemia (but not hypercholesterolemia) and fatty infiltrations of NAFLD (26), emphasizing the importance of separately analyzing the genders in order not to mask potential correlates that are gender dependent. Given the high rates of end-stage liver disease in our previous study where we found significantly higher rates in males than females (1), this gender difference may point to important metabolic pathways and/or behavioral differences in this population that lead to different rates of NAFLD in each gender (27,28). In addition, our analysis highlighted the critical role of non-HDL-c rather than total cholesterol in liver injury. A very minor risk effect of total cholesterol could be identified in our modeling provided we did not discriminate non-HDL-c from total cholesterol, because HDL-c counteracted and diluted the risk of non-HDL-c. On the other hand, LDL-c calculated according to the formula of Friedewald et al. (29) is not associated with increased ALT in our study. This finding provides additional evidence that, instead of calculated LDL-c, non-HDL-c is a risk marker of metabolic syndrome (30) or the liver injury of metabolic syndrome NAFLD in our population. The machine learning approach adds to our previous observations by identifying diastolic hypertension as an independent major risk factor for liver injury in this Hispanic community. Although diastolic hypertension is a common complication of the metabolic syndrome, it is not known whether there is an independent pathogenic mechanism associating it with NAFLD.

A gender-specific genetic susceptibility was highlighted in this study. As shown by our study, the genetic susceptibility tagged by the *PNPLA3* SNP rs738409 is only seen in females. The lack of association in males cannot be explained by sample size or statistical power. Comparing the genetic effect in males (OR [95% CI] = 1.171 [0.911, 1.505]) and females (OR [95% CI] = 1.640 [1.347, 1.996]), the heterogeneity is statistical significance (p = 0.038). In addition to our study, a similar gender-specific effect has been reported by meta-analysis of genetic association of rs738409 and NAFLD (31). Molecular mechanisms underlying this gender-specific effect remain unknown, which is being investigated in our future study. In spite of the highly significant association, the role of rs738409 in the prediction model is minimal compared to other epidemiological risk factors.

We found that history of hepatitis contributed only a minor risk for elevated ALT in males. Our previous unpublished data show very low rates of hepatitis C seropositivity (0/320) and hepatitis B (3/320) in randomly collected sera from this population (Fisher-Hoch, unpublished data). We also found that aging in itself was not a risk factor for NAFLD. On the contrary, evidence of liver injury was most marked in younger males who are also more likely to be severely obese (5). Neither drugs nor alcohol consumption was an indicator of liver injury in this study; however, a history of medication (including any prescription drugs by a physician) did show a protective effect in males. This may be correlated with receiving treatment for insulin resistance, hypertriglyceridemia, or hypertension. Although physical activity showed a trend towards a protective effect, it did not reach statistical significance, perhaps due to imprecise quantification of physical activity in this study. The interesting finding of this study that FBG does not contribute to the high risk of elevated ALT in males highlighted that intensive control of blood glucose alone may not be a viable therapeutic target for liver injury in metabolic syndrome.

We identified and stratified the most important and specific risk factors for liver injury in this cohort of Mexican-American subjects. Our conclusions are consistent with previous studies on subjects with a clear diagnosis of NAFLD (17,18) because risk factors for liver injury in our study are largely concordant with previous studies on NAFLD. In a group of Korean subjects with NAFLD, Oh et al. showed that increased ALT was associated with serum lipid (increased triglycerides and non-HDL-c, and decreased HDL-c), insulin resistance (fasting glucose, fasting insulin, and HOMA-IR), body fat (waist circumference and BMI), and hypertension (diastolic blood pressure) (17). In a group of Italian subjects with NAFLD, Bedogni et al. showed increased risk of NAFLD due to obesity, hyperglycemia, hypertriglyceridemia, and systolic hypertension (18). In comparison, the distribution pattern of risk factors identified by our study suggests NAFLD as likely the most important cause of increased ALT in Mexican-Americans. Our model allows us to identify NAFLD as the major cause of liver injury in this Mexican-American population. Globally, NAFLD is a fast emerging disease and has recently received extensive attention.

The definitive diagnosis of NAFLD is difficult. Liver biopsy is the “gold standard” to diagnose NAFLD and to differentiate it from NASH because it determines the presence and extent of hepatic fibrosis (32). However, because of the potential serious complications (0.1% major hemorrhage, and 0.01% death) and its technical complexity, inconvenience to the participant and cost, liver biopsy is not suitable for use in a screening study at the population level (33) so was not considered appropriate in this study. Despite these limitations, our data identify NAFLD as a likely major contributor to the extraordinarily high rate of liver injury in this Mexican-American population with limited access to health care. Community-based efforts and simple preventive medicine to reduce NAFLD are critically needed to significantly lower the burden of liver injury, including end-stage liver disease, in disadvantaged populations.

Acknowledgments

We thank our cohort recruitment team, particularly Rocio Uribe, Elizabeth Braunstein and Julie Ramirez. We also thank Marcela Montemayor and other laboratory staff for their contribution and Christina Villarreal for administrative support. We thank Valley Baptist Medical Center, Brownsville for providing us space for our Center for Clinical and Translational Science Clinical Research Unit. We also thank the community of Brownsville and the participants who so willingly participated in this study in their city.

This work was supported by MD000170 P20 funded from the National Center on Minority Health and Health Disparities (NCMHD) and the Centers for Translational Science Award 1U54RR023417–01 from the National Center for Research Resources (NCRR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. HQQ is supported by intramural funding from the University of Texas, School of Public Health.

References

1. Perez A, Anzaldua M, McCormick J, et al. High frequency of chronic end-stage liver disease and hepatocellular carcinoma in a Hispanic population. J Gastroenterol Hepatol. 2004; 19:289–295. [PubMed: 14748876]
2. Fisher-Hoch SP, Rentfro AR, Salinas JJ, et al. Socioeconomic status and prevalence of obesity and diabetes in a Mexican American community, Cameron County, Texas, 2004–2007. Prev Chronic Dis. 2010; 7:A53. [PubMed: 20394692]
3. Qu HQ, Li Q, Rentfro AR, et al. The definition of insulin resistance using HOMA-IR for Americans of Mexican descent using machine learning. PLoS One. 2011; 6:e21041. [PubMed: 21695082]
4. Pan JJ, Qu HQ, Rentfro A, et al. Prevalence of metabolic syndrome and risks of abnormal serum alanine aminotransferase in Hispanics: a population-based study. PLoS One. 2011; 6:e21515. [PubMed: 21720553]
5. Salinas J, McCormick JB, Rentfro A, et al. The missing men: high risk of disease in men of Mexican origin. Am J Mens Health. 2011; 5:332–340. [PubMed: 20930218]
6. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2010; 33:S62–S69. [PubMed: 20042775]
7. Romeo S, Kozlitina J, Xing C, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet. 2008; 40:1461–1465. [PubMed: 18820647]

8. Chalasani N, Guo X, Loomba R, et al. Genome-wide association study identifies variants associated with histologic features of nonalcoholic fatty liver disease. Gastroenterology. 2010; 139:1567–1576. 1576 e1561–1566. [PubMed: 20708005]
9. Yuan X, Waterworth D, Perry JRB, et al. Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. Am J Human Genet. 2008; 83:520–528. [PubMed: 18940312]
10. Cortes C, Vapnik V. Support-vector networks. Machine Learning. 1995; 20:273–297.
11. Noble WS. What is a support vector machine? Nat Biotechnol. 2006; 24:1565–1567. [PubMed: 17160063]
12. Wei Z, Wang K, Qu HQ, et al. From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. PLoS Genet. 2009; 5:e1000678. [PubMed: 19816555]
13. Clark TG, De Iorio M, Griffiths RC. Bayesian logistic regression using a perfect phylogeny. Biostatistics. 2007; 8:32–52. [PubMed: 16556611]
14. Matthews DR, Hosker JP, Rudenski AS, et al. Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985; 28:412–419. [PubMed: 3899825]
15. Fan R-E, Chen P-H, Lin CJ. Working set selection using second order information for training support vector machines. J Mach Learn Res. 2005; 6:1889–1918.
16. Chen, YW.; Lin, CJ. Feature Extraction, Foundations and Applications. New York: Springer; 2006. Combining SVMs with various feature selection strategies.
17. Oh SY, Cho YK, Kang MS, et al. The association between increased alanine aminotransferase activity and metabolic factors in nonalcoholic fatty liver disease. Metabolism. 2006; 55:1604–1609. [PubMed: 17142131]
18. Bedogni G, Miglioli L, Masutti F, et al. Prevalence of and risk factors for nonalcoholic fatty liver disease: the Dionysos nutrition and liver study. Hepatology. 2005; 42:44–52. [PubMed: 15895401]
19. Falck-Ytter Y, Younossi ZM, Marchesini G, et al. Clinical features and natural history of nonalcoholic steatosis syndromes. Semin Liver Dis. 2001; 21:017–026.
20. Tilg H. Adipocytokines in nonalcoholic fatty liver disease: key players regulating steatosis, inflammation and fibrosis. Curr Pharm Des. 2010; 16:1893–1895. [PubMed: 20370678]
21. Sanyal D, Mukhopadhyay P, Pandit K, et al. Central obesity but not generalised obesity (body mass index) predicts high prevalence of fatty liver (NRFLD), in recently detected untreated, IGT and type 2 diabetes Indian subjects. J Indian Med Assoc. 2009; 107:755–758. [PubMed: 20469778]
22. Polyzos SA, Kountouras J, Zavos C. Nonalcoholic fatty liver disease: the pathogenetic roles of insulin resistance and adipocytokines. Curr Mol Med. 2009; 9:299–314. [PubMed: 19355912]
23. Vanni E, Bugianesi E. The gut-liver axis in nonalcoholic fatty liver disease: another pathway to insulin resistance? Hepatology. 2009; 49:1790–1792. [PubMed: 19475679]
24. Shoelson SE, Lee J, Goldfine AB. Inflammation and insulin resistance. J Clin Investig. 2006; 116:1793–1801. [PubMed: 16823477]
25. Hotamisligil GS. Inflammation and metabolic disorders. Nature. 2006; 444:860–867. [PubMed: 17167474]
26. Assy N, Kaita K, Mymin D, et al. Fatty infiltration of liver in hyperlipidemic patients. Dig Dis Sci. 2000; 45:1929–1934. [PubMed: 11117562]
27. Hashimoto E, Tokushige K. Prevalence, gender, ethnic variations, and prognosis of NASH. J Gastroenterol. 2011; 46(suppl 1):S63–S69.
28. Christensen KE, Wu Q, Wang X, et al. Steatosis in mice is associated with gender, folate intake, and expression of genes of one-carbon metabolism. J Nutr. 2010; 140:1736–1741. [PubMed: 20724492]
29. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972; 18:499–502. [PubMed: 4337382]

30. Arsenault BJ, Boekholdt SM, Kastelein JJ. Lipid parameters for measuring risk of cardiovascular disease. Nat Rev Cardiol. 2011; 8:197–206. [PubMed: 21283149]

31. Sookoian S, Pirola CJ. Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. Hepatology. 2011; 53:1883–1894. [PubMed: 21381068]

32. Myers RP. Noninvasive diagnosis of nonalcoholic fatty liver disease. Ann Hepatol. 2009; 8(suppl 1):S25–S33. [PubMed: 19381121]

33. Bravo AA, Sheth SG, Chopra S. Liver biopsy. N Engl J Med. 2001; 344:495–500. [PubMed: 11172192]

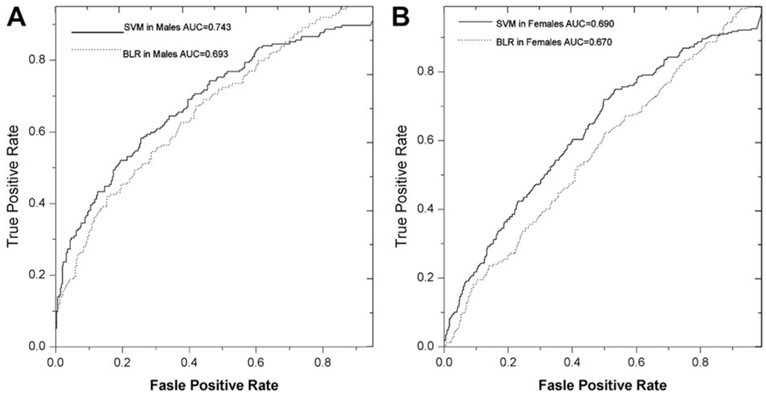

Figure 1.

The risk modeling of liver injury in the Cameron Cohort Hispanic Cohort (CCHC). (A) SVM and BLR models in males. (B) SVM and BLR model in females. As shown by the AUROC score, the SVM model has better performance in both males and females than the BLR model.

Table 1

Risk model of liver injury in the Cameron cohort (phenotypic mean ± standard deviation)

	Males					Females				
Variable	SVM F score[a]	BLR betavalue[b]	Normal ALT group (n = 221[d], $\bar{X}$ ± s)	Elevated ALT group (n = 274[d], $\bar{X}$ ± s)	p	SVM F score[a]	BLR beta value[b]	Normal ALT group (n = 752[d], $\bar{X}$ ± s)	Elevated ALT group (n = 282[d], $\bar{X}$ ± s)	p
Serum insulin level (U/L)	0.081	2.481	12.0 ± 8.6	19.5 ± 15.1	6.22×10^{-11}	0.075	3.681	12.5 ± 7.6	19.8 ± 13.0	8.48×10^{-27}
BMI	0.073	1.711	28.5 ± 7.7	31.7 ± 6.4	5.40×10^{-7}	0.043	0.699	29.8 ± 7.1	33.6 ± 7.8	2.18×10^{-13}
Age	0.065	−1.986	51.6 ± 20.7	40.6 ± 15.1	4.08×10^{-11}	0.002	−1.391	48.7 ± 19.5	45.1 ± 15.9	6.38×10^{-3}
Body weight (kg)	0.059	0.506	83.8 ± 22.8	92.6 ± 21.0	9.81×10^{-6}	0.042	1.229	72.3 ± 17.7	80.4 ± 18.6	1.84×10^{-10}
Education level (years)	0.046	1.076	10.1 ± 4.5	11.1 ± 3.5	3.30×10^{-3}	0.000	0.326	9.5 ± 4.4	9.7 ± 4.0	0.514
Non-HDL-c (mg/dL)	0.042	2.347	138.3 ± 41.1	153.4 ± 37.7	2.61×10^{-5}	0.006	1.403	136.7 ± 37.2	144.6 ± 40.5	3.16×10^{-3}
HOMA-IR	0.037	0.436	3.6 ± 3.5	5.3 ± 4.6	6.10×10^{-6}	0.079	2.667	3.3 ± 2.4	6.3 ± 5.8	1.52×10^{-29}
DBP (mmHg)	0.037	1.129	71.6 ± 10.5	74.8 ± 10.0	6.91×10^{-4}	0.017	2.009	68.5 ± 9.7	71.0 ± 10.0	2.64×10^{-4}
Hip (cm)	0.034	0.085	107.6 ± 13.7	111.2 ± 10.8	1.39×10^{-3}	0.027	−0.589	108.6 ± 15.1	114.0 ± 13.2	1.99×10^{-7}
Waist (cm)	0.033	0.917	101.8 ± 16.4	106.4 ± 14.6	1.17×10^{-3}	0.038	1.400	97.1 ± 16.2	104.5 ± 14.5	4.38×10^{-11}
Serum triglycerides (mg/dL)	0.025	1.996	157.0 ± 91.0	222.2 ± 295.2	1.11×10^{-3}	0.019	1.384	144.2 ± 93.2	204.3 ± 173.5	2.05×10^{-12}
Medications (no vs. yes)	0.025	−0.157	146 vs. 96	191 vs. 62	1.44×10^{-3}[c]	0.001	0.015	455 vs. 303	183 vs. 91	0.138[c]
Waist/hip	0.019	1.826	0.9 ± 0.1	1.0 ± 0.1	9.18×10^{-3}	0.011	0.236	0.9 ± 0.1	0.9 ± 0.1	1.73×10^{-5}
Annual income (USD)	0.016	0.557	19064.0 ± 15626.8	23742.6 ± 25861.6	0.016	0.000	−0.745	16068.0 ± 15623.7	14525.0 ± 13129.2	0.145
Hepatitis history (no vs. yes)	0.015	1.813	241 vs. 0	243 vs. 10	0.002[c]	0.001	1.120	741 vs. 18	266 vs. 8	0.884[c]
FPG (mg/dL)	0.012	−1.670	117.3 ± 55.5	109.3 ± 39.9	0.065	0.017	0.740	105.0 ± 36.0	123.7 ± 66.4	1.04×10^{-8}
Smoking (no vs. yes)	0.009	−0.678	229 vs. 13	247 vs. 7	0.139[c]	0.001	−0.309	755 vs. 4	273 vs. 2	0.708[c]
Diabetes diagnosis (nondiabetic, prediabetic, diabetic)	0.008	−0.146	85, 68, 89	86, 100, 68	0.013	0.007	−0.058	339, 169, 251	101, 85, 88	0.010
Height (cm)	0.008	0.495	167.8 ± 23.3	170.5 ± 6.8	0.074	0.002	0.277	154.9 ± 11.1	154.7 ± 6.1	0.802
Heavy physical activity (h/week)	0.007	0.078	1.5 ± 2.7	1.9 ± 3.0	0.128	0.000	−0.436	0.5 ± 1.3	0.4 ± 1.3	0.464
Light physical activity (h/week)	0.004	−1.431	1.2 ± 1.4	1.3 ± 1.6	0.606	0.002	−1.841	1.5 ± 1.6	1.4 ± 1.5	0.127
rs738409 (G/G, G/C, C/C)	0.003	0.390	54, 137, 51	54, 142, 57	0.915[c]	0.024	1.179	218, 404, 138	36, 136, 103	3.37×10^{-12}[c]
HDL-c (mg/dL)	0.001	−0.310	44.1 ± 10.9	43.1 ± 9.3	0.229	0.008	−0.590	50.8 ± 12.0	46.7 ± 11.3	9.07×10^{-7}
Pulse rate (per min)	0.001	−0.378	33.2 ± 5.4	33.5 ± 4.9	0.518	0.001	−0.383	34.8 ± 4.9	34.5 ± 5.1	0.425
SBP (mmHg)	0.001	−0.561	121.2 ± 16.1	118.6 ± 13.6	0.057	0.001	−0.922	116.1 ± 18.3	117.7 ± 18.3	0.215

	Males					Females				
Variable	**SVM F score**[a]	**BLR betavalue**[b]	**Normal ALT group (*n* = 221**[d]**, $\bar{X} \pm s$)**	**Elevated ALT group (*n* = 274**[d]**, $\bar{X} \pm s$)**	***p***	**SVM F score**[a]	**BLR beta value**[b]	**Normal ALT group (*n* = 752**[d]**, $\bar{X} \pm s$)**	**Elevated ALT group (*n* = 282**[d]**, $\bar{X} \pm s$)**	***p***
LDL-c (mg/dL)	0.001	−0.579	106.1 ± 38.5	109.9 ± 41.4	0.285	0.000	−0.198	106.1 ± 35.0	98.1 ± 47.0	3.57×10^{-3}
Moderate physical activity (h/week)	0.000	−0.079	7.5 ± 3.5	7.6 ± 3.6	0.740	0.001	−0.712	8.4 ± 4.0	9.0 ± 3.2	0.02
Alcohol consumption (no vs. yes)	0.000	−0.009	58 vs. 184	64 vs. 190	0.751[c]	0.000	−0.015	490 vs. 270	195 vs. 79	0.045[c]

ALT, alanine aminotransferase; BLR, Bayesian logistic regression; SBP, systolic blood pressure; DBP, diastolic blood pressure; HDL-c, high-density lipoprotein cholesterol; LDL-c, low-density lipoprotein cholesterol; FPG, fasting plasma glucose; BMI, body mass index.

[a] F-score measures the discrimination of two groups, whereas larger F-score suggests the feature is more discriminative by the variable.

[b] BLR beta value is the maximum likelihood values of the slope parameters in the BLR model, whereas larger absolute value suggests larger effect by the factor, and +/− represents the direction of the correlation.

[c] χ^2 test *p* value.

[d] Actual unweighted numbers investigated in this study.