Cell Reports Medicine

CellPress
OPEN ACCESS

Preview

The quest for the missing links in fatty liver genetics: Deep learning to the rescue!

Sami Qadri[1,2] and Hannele Yki-Järvinen[1,2,*]
[1]University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[2]Minerva Foundation Institute for Medical Research, Helsinki, Finland
*Correspondence: hannele.yki-jarvinen@helsinki.fi
https://doi.org/10.1016/j.xcrm.2022.100862

Park, MacLean, et al. conduct an exome-wide association study of liver fat content in the Penn Medicine BioBank.[1] By leveraging machine learning-assisted analysis of clinical CT scans to quantify steatosis, they uncover previously undescribed liver fat-associated genetic variants.

Non-alcoholic fatty liver disease (NAFLD) is the most common chronic liver disease worldwide. Encompassing a spectrum of conditions, NAFLD ranges from simple steatosis (increased liver triglycerides without necroinflammation or fibrosis) to non-alcoholic steatohepatitis and cirrhosis. Patients with NAFLD frequently embody features of the insulin resistance/metabolic syndrome including obesity, hypertension, dyslipidemia, and type 2 diabetes, and they are therefore at increased risk of cardiovascular disease.[2] In the United States, end-stage liver disease due to NAFLD is already the second leading indication for liver transplantation, accounting for approximately one-fourth of all cases.[3]

Not all patients with NAFLD are alike, however. Genetic studies have been highly successful and have identified common variants in genes such as *PNPLA3*, *TM6SF2*, *MBOAT7*, *MTARC1*, *GPAM*, and *APOE*, which together explain >20% of the population attributable risk of steatosis and >30% of that of cirrhosis.[4] Importantly, the causes and consequences of NAFLD due to genetic polymorphisms clearly differ from those underlying NAFLD associated with insulin resistance.[5] Insulin-resistant patients (and their livers) practically bathe in a substrate surplus, which, in conjunction with hyperinsulinemia, fuels hepatic triglyceride synthesis via multiple pathways.[5] This is not the case in NAFLD due to a high burden of liver fat-increasing genetic variants, the exact mechanisms of which are unclear but likely relate to specific defects in hepatic fatty acid export.[5] It is also of clinical interest that variants in *PNPLA3* and *TM6SF2* predispose to all stages of NAFLD but paradoxically *protect* against atherogenic dyslipidemia[6] and cardiovascular disease.[7]

Despite these recent advances in understanding the molecular genetic underpinnings of NAFLD, no approved pharmaceuticals to treat the disease or its progressive forms exist. Genetic discovery studies have the potential of unearthing novel disease-related proteins, which may be amenable to modulation via precision medicine approaches.

In this issue, Park, MacLean, et al. conducted an exome-wide association study of liver fat content in 10,283 participants of the Penn Medicine BioBank (PMBB).[1] To quantify liver fat in clinical CT scans of the chest and abdomen, they developed and validated a fully automated image curation and organ labeling technique using deep learning. This involved training of three separate neural networks to first identify non-contrast CT scans, followed by segmentation of the liver and spleen in axial 2D slices for calculation of the difference in attenuation between the organs. This spleen-liver attenuation difference is an established albeit insensitive and semi-quantitative biomarker of liver fat, which is based on the observation that increased liver fat leads to decreased hepatic attenuation in CT imaging.[8] Here, the automatically calculated attenuation difference associated not only with the diagnosis of chronic liver disease and increased liver enzymes but also with a host of other cardiometabolic conditions and related laboratory findings[1]—as would be expected for a *bona fide* biomarker of liver fat. Thus, it served as a proxy for steatosis in the exome-wide analysis.

Among 120,315 total variants tested, the primary association analysis of single coding variants identified 91 variants in 86 genes with exome-wide significant or suggestive associations with liver fat.[1] This was followed by replication in the UK Biobank (UKB) in which 9,049 individuals had undergone whole-exome sequencing and liver fat measurement by magnetic resonance imaging. In summary, the study confirmed several variants previously associated with steatosis (*PNPLA3* I148M, *TM6SF2* E167K, *APOE* C130R) and identified two additional, heretofore undescribed variants (*FGD5* H600Y and *CITED2* S198_G199del) that also replicated in the UKB. In phenome-wide analysis, the two novel variants associated with several hepatic and dysmetabolism-related outcomes. Furthermore, a gene burden of rare predicted loss-of-function variants in *LMF2* significantly associated with liver fat in both the PMBB and UKB cohorts.

Deep learning is a popular machine learning paradigm based on artificial neural networks that is well-suited for various image recognition and classification tasks. Usual applications in medical image analysis include automatic labeling of images (e.g., detection of non-contrast vs. contrast-enhanced CT scans), segmentation of objects of interest within

an image (delineation of the liver and spleen in axial CT slices), or detecting and counting of objects such as specific cell types, nuclei, or mitoses in histological specimens.[9] Relevant to the present study by Park, MacLean, et al., the capacity of a computer to process vast amounts of image data per unit time while upholding a consistent quality of analysis far surpasses that of humans. Thus, deep learning can be harnessed to greatly accelerate a project entailing interpretation of images, potentially reducing the time to completion from years or months to weeks or even days. Here, the authors not only identified an under-utilized imaging dataset but they ingeniously applied deep learning to first filter out contrast-enhanced images and then derive the final liver fat measurements. Similar to the present study, several investigators have leveraged deep learning to quantify liver fat in clinical CT scans (reviewed by Starekova et al.[8]). To our knowledge, however, this study is the first to use such data in a genome-wide association study of NAFLD. It deserves a mention that Speliotes and co-workers have conducted similar analyses in large cohorts with CT-derived liver fat measurements, but without using computerized assistance.[10]

As in previous studies, the association of *PNPLA3* I148M with liver fat was by far the most significant, with a p value of 1.91E−30.[1] The p value for the *TM6SF2* E167K variant was 1.74E−15. Most of the other variants, including the two novel ones (*FGD5* H600Y and *CITED2* S198_G199del), had a p value ranging on the order of magnitude from E−4 to E−6, i.e., up to septillion times higher than that of *PNPLA3* I148M. As no histological validation or functional studies were performed with the replicated variants, their significance in regulating liver fat content remains uncertain. Nevertheless, the authors should be congratulated for being able to extract essentially new data out of the existing PMBB imaging dataset—without the need to restudy the patients—and successfully combine these data with exome sequencing to identify potentially novel NAFLD-associated genes worth investigating. These types of studies emphasize the huge potential of deep learning-based techniques combined with clinical and genetic information in unraveling the mysteries of human disease.

DECLARATION OF INTERESTS

S.Q. declares no competing interests. H.Y.-J. serves as a consultant to Novo Nordisk, GSK, and Inipharm. In addition, H.Y.-J. is an advisory board member of MSD, Eli Lilly, and Hanmi Pharmaceutical.

REFERENCES

1. Park, J., MacLean, M.T., Lucas, A.M., Torigian, D.A., Schneider, C.V., Cherlin, T., Xiao, B., Miller, J.E., Bradford, Y., Judy, R.L., et al. (2022). Exome-wide analysis of hepatic fat derived from clinical CT imaging in a medical biobank. Cell Rep. Med. *3*, 100855.
2. Yki-Jarvinen, H. (2014). Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. Lancet Diabetes Endocrinol. *2*, 901–910. https://doi.org/10.1016/S2213-8587(14)70032-4.
3. Younossi, Z.M., Stepanova, M., Ong, J., Trimble, G., AlQahtani, S., Younossi, I., Ahmed, A., Racila, A., and Henry, L. (2021). Nonalcoholic steatohepatitis is the most Rapidly increasing indication for liver transplantation in the United States. Clin. Gastroenterol. Hepatol. *19*, 580–589.e5. https://doi.org/10.1016/j.cgh.2020.05.064.
4. Jamialahmadi, O., Mancina, R.M., Ciociola, E., Tavaglione, F., Luukkonen, P.K., Baselli, G., Malvestiti, F., Thuillier, D., Raverdy, V., Mannisto, V., et al. (2021). Exome-wide association study on Alanine Aminotransferase Identifies sequence variants in the GPAM and APOE associated with fatty liver disease. Gastroenterology *160*, 1634–1646.e7. https://doi.org/10.1053/j.gastro.2020.12.023.
5. Luukkonen, P.K., Qadri, S., Ahlholm, N., Porthan, K., Mannisto, V., Sammalkorpi, H., Penttila, A.K., Hakkarainen, A., Lehtimaki, T.E., Gaggini, M., et al. (2022). Distinct contributions of metabolic dysfunction and genetic risk factors in the pathogenesis of non-alcoholic fatty liver disease. J. Hepatol. *76*, 526–535. https://doi.org/10.1016/j.jhep.2021.10.013.
6. Luukkonen, P.K., Qadri, S., Lehtimaki, T.E., Juuti, A., Sammalkorpi, H., Penttila, A.K., Hakkarainen, A., Orho-Melander, M., Arola, J., and Yki-Jarvinen, H. (2021). The PNPLA3-I148M variant Confers an Antiatherogenic Lipid Profile in insulin-resistant patients. J. Clin. Endocrinol. Metab. *106*, e300–e315. https://doi.org/10.1210/clinem/dgaa729.
7. Liu, D.J., Peloso, G.M., Yu, H., Butterworth, A.S., Wang, X., Mahajan, A., Saleheen, D., Emdin, C., Alam, D., Alves, A.C., et al. (2017). Exome-wide association study of plasma lipids in >300,000 individuals. Nat. Genet. *49*, 1758–1766. https://doi.org/10.1038/ng.3977.
8. Starekova, J., Hernando, D., Pickhardt, P.J., and Reeder, S.B. (2021). Quantification of liver fat content with CT and MRI: State of the Art. Radiology *301*, 250–262. https://doi.org/10.1148/radiol.2021204288.
9. Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., Liu, Y., Topol, E., Dean, J., and Socher, R. (2021). Deep learning-enabled medical computer vision. NPJ Digit Med *4*, 5. https://doi.org/10.1038/s41746-020-00376-2.
10. Palmer, N.D., Kahali, B., Kuppa, A., Chen, Y., Du, X., Feitosa, M.F., Bielak, L.F., O'Connell, J.R., Musani, S.K., Guo, X., et al. (2021). Allele-specific variation at APOE increases nonalcoholic fatty liver disease and obesity but decreases risk of Alzheimer's disease and myocardial infarction. Hum. Mol. Genet. *30*, 1443–1456. https://doi.org/10.1093/hmg/ddab096.