

ORIGINAL ARTICLE

OPEN

A functional genomic framework to elucidate novel causal metabolic dysfunction–associated fatty liver disease genes

Peter Saliba-Gustafsson^{1,2,3,4}  | Johanne M. Justesen^{1,5}  |
 Amanda Ranta^{1,3,4}  | Disha Sharma^{1,4}  | Ewa Bielczyk-Maczynska^{1,3,4,6}  |
 Jiehan Li^{1,3,4} | Laeya A. Najmi^{1,3,4} | Maider Apodaka⁷  |
 Patricia Aspichueta^{7,8}  | Hanna M. Björck⁹  | Per Eriksson⁹  |
 Theresa M. Schurr^{1,5}  | Anders Franco-Cereceda¹⁰  | Mike Gloudemans¹¹  |
 Endrina Mujica¹²  | Marcel den Hoed¹²  | Themistocles L. Assimes^{1,13}  |
 Thomas Quertermous^{1,5}  | Ivan Carcamo-Orive^{1,14,15,16}  | Chong Y. Park¹  |
 Joshua W. Knowles^{1,3,4,17} 

¹Department of Medicine, Division of Cardiovascular Medicine and Cardiovascular Institute, Stanford University, Stanford, California, USA

²CardioMetabolic Unit, Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden

³Department of Medicine, Stanford Diabetes Research Center, Stanford University, California, USA

⁴Stanford Cardiovascular Institute, Stanford School of Medicine, Stanford University, California, USA

⁵Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark

⁶The Hormel Institute, University of Minnesota, Minneapolis, Minnesota, USA

⁷Department of Physiology, University of the Basque Country (UPV/EHU), Faculty of Medicine and Nursing, Leioa, Spain

⁸National Institute for the Study of Liver and Gastrointestinal Diseases (CIBERehd, Instituto de Salud Carlos III)

⁹Department of Medicine, Division of Cardiovascular Medicine, Centre for Molecular Medicine, Karolinska Institutet, Stockholm, Karolinska University Hospital, Solna, Sweden

¹⁰Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden

¹¹Department of Pathology, Stanford University School of Medicine, CA, USA

¹²Department of Immunology, Genetics and Pathology, Uppsala University, Sweden

¹³VA Palo Alto Health Care System, Palo Alto, California, USA

¹⁴Department of Endocrinology, Metabolism, Nutrition, and Kidney Disease, Biobizkaia Health Research Institute, Cruces, Spain

¹⁵IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

¹⁶Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU), Leioa, Spain

¹⁷Stanford Prevention Research Center, Stanford, California, USA

Abbreviations: BMI, body mass index; CRISPRi, CRISPR-interference; FLI, fatty liver index; GWAS, genome-wide association studies; MASLD, metabolic dysfunction–associated fatty liver disease; MASLD-S, metabolic dysfunction–associated fatty liver disease-score; QTL, quantitative trait locus; scRNA, single-cell RNA sequencing; sgRNA, small guide RNA; UKB, UK Biobank.

Peter Saliba-Gustafsson and Johanne M. Justesen contributed equally to this work.

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, www.hepjournal.com.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Copyright © 2024 The Author(s). Published by Wolters Kluwer Health, Inc.

Correspondence

Joshua W. Knowles, Falk Cardiovascular Research Center, Stanford University, Stanford School of Medicine, Cardiovascular Medicine, 870 Quarry Road Extension, Stanford, CA 94305, USA.
Email: knowlej@stanford.edu

Peter Saliba-Gustafsson, C2:94, Karolinska Universitetssjukhuset Huddinge, 14186 Stockholm.
Email: peter.gustafsson.1@ki.se

Ivan Carcamo-Orive, BioBizkaia Health Research Institute, Pza. de Cruces, S/N. Edificio Biobizkaia 1, 48903, Barakaldo, Bizkaia, Spain.
Email: icarcamo.ikerbasque@gmail.com

Abstract

Background and Aims: Metabolic dysfunction–associated fatty liver disease (MASLD) is the most prevalent chronic liver pathology in western countries, with serious public health consequences. Efforts to identify causal genes for MASLD have been hampered by the relative paucity of human data from gold standard magnetic resonance quantification of hepatic fat. To overcome insufficient sample size, genome-wide association studies using MASLD surrogate phenotypes have been used, but only a small number of loci have been identified to date. In this study, we combined genome-wide association studies of MASLD composite surrogate phenotypes with genetic colocalization studies followed by functional in vitro screens to identify bona fide causal genes for MASLD.

Approach and Results: We used the UK Biobank to explore the associations of our novel MASLD score, and genetic colocalization to prioritize putative causal genes for in vitro validation. We created a functional genomic framework to study MASLD genes in vitro using CRISPRi. Our data identify *VKORC1*, *TNKS*, *LYPLAL1*, and *GPAM* as regulators of lipid accumulation in hepatocytes and suggest the involvement of *VKORC1* in the lipid storage related to the development of MASLD.

Conclusions: Complementary genetic and genomic approaches are useful for the identification of MASLD genes. Our data supports *VKORC1* as a bona fide MASLD gene. We have established a functional genomic framework to study at scale putative novel MASLD genes from human genetic association studies.

INTRODUCTION

Metabolic dysfunction–associated fatty liver disease (MASLD) is the most common chronic liver condition, with serious public health consequences. Globally, at least 25% of adults are estimated to suffer from MASLD, and cardiovascular disease is the leading cause of death among these patients.^[1,2] MASLD displays a wide spectrum of liver pathology, ranging from NAFL, which is typically benign, to metabolic dysfunction–associated steatohepatitis, characterized by steatosis and features of cellular injury, such as inflammation and hepatocyte ballooning. Metabolic dysfunction–associated steatohepatitis may progress to liver cirrhosis, hepatic failure, and HCC in the absence of significant alcohol consumption. The degree of steatosis can be measured through various imaging techniques but the gold standard of these is abdominal MRI. However, abdominal MRI is not typically conducted on asymptomatic individuals, often leaving MASLD undiagnosed for years.

Genome-wide association studies (GWAS) have been used to identify associations between MASLD and common genetic variants.^[3–5] Due to the scarcity of MRI data, identifying risk loci for MASLD has been slower than

for other cardiometabolic diseases or their risk factors (eg, body mass index (BMI) or biochemical measures (eg, serum liver enzymes and lipids levels), and other complex cardiometabolic diseases such as obesity, and diabetes. One way to overcome data scarcity in MASLD is to comprise latent proxies for MASLD using data more readily available in large cohort studies. For instance, Bedogni et al^[6] established the fatty liver index (FLI) as a surrogate variable for MASLD; however, FLI did not outperform waist circumference in predicting MASLD in a validation study.^[7] Recently, Haas et al^[8] used the then-largest available data set on MRI-derived MASLD, together with machine learning, to identify five new MASLD risk loci. Nevertheless, significant gaps remain in the understanding of the genetic architecture of MASLD.

Aiming to increase our understanding of the molecular etiology of MASLD, we here generate (MASLD-S), a composite variable of anthropometric and biochemical variables to predict liver fat. By using an alternative surrogate to predict liver fat, and running GWAS combined with genetic colocalization, we identify novel loci associated with MASLD. We use CRISPR-interference (CRISPRi) to interrogate the impact of multiple genes on both transcriptional changes and functional phenotypes, at

TABLE 1 Variables used to construct the MASLD-S with their respective estimates, SEs, Z-statistics, and *p*-values

Variable	Estimate	SE	Z Stat	<i>p</i>
Intercept	-3.20145	0.34775	-9.206	2.00E-16
Sex	-0.52707	0.0803	-6.564	5.23E-11
Waist Circ	0.53716	0.06679	8.043	8.79E-16
GGT	0.06514	0.03899	1.671	0.0948
BMI	0.06973	0.01178	5.921	3.20E-09
Cholesterol	-0.14647	0.03079	-4.757	1.97E-06
AST	0.15941	0.03241	4.918	8.73E-07
HbA1c	0.15286	0.03031	5.043	4.58E-07
AST/ALT	-0.40867	0.03461	-11.809	2.00E-16
TGs	0.45029	0.03574	12.598	2.00E-16
Alb	0.17569	0.03024	5.811	6.23E-09

Abbreviations: Alb, albumin; BMI, body mass index; MASLD-S, metabolic dysfunction-associated fatty liver disease-score; TGs, triglycerides.

a single-cell level.^[9-11] We characterize a subset of putative MASLD genes *in vitro* and *in vivo* through an integrated framework and identify *VKORC1* as a likely causal MASLD gene.

EXPERIMENTAL PROCEDURES

For details, please refer to the Supplemental Materials, <http://links.lww.com/HEP/I623>.

Human molecular genetics

This research was conducted using a purposeful subset of the UK Biobank (UKB) data under application number 13721, described in detail in the Extended Methods. Research was conducted in accordance with both the Declarations of Helsinki and Istanbul. Ethical approval was obtained from the Swedish Ethical Review Authority (approval nr 2006/784-31/1). A new surrogate marker for MASLD (MASLD score) was generated, whose predictive value was compared to other biochemical surrogate markers of MASLD. Human molecular genetic approaches, including genetic colocalization, were adopted to select target genes for functional genomic experimentation. The colocalization analyses were carried out using custom pipelines, described in detail in the Extended Methods, found in the Supplemental Material, <http://links.lww.com/HEP/I623>.

CRISPRi and perturb-seq

A human hepatocyte cell line, suitable for large-scale CRISPRi and Perturb-seq, was generated and characterized using single-cell RNA-seq. Details are described

in the Extended Methods, found in the Supplemental Material, <http://links.lww.com/HEP/I623>.

CRISPRi for hepatocyte lipid accumulation and Perturb-seq was conducted for selected target genes in the HepaRG cell line under established culturing protocols. The *VKORC1* gene was validated as a novel MASLD gene using single small guide RNA (sgRNA) knockdowns. Transcript levels of *VKORC1* and *PLIN2* were analyzed in HepaRG cells on gene knockdown. To functionally assess the role of *VKORC1* in MASLD, we explored lipid droplet staining in HepaRG cells where *VKORC1* had been knocked down, described in detail in the Extended Methods, found in the Supplemental Material, <http://links.lww.com/HEP/I623>.

Validation of *VKORC1* as a MASLD gene in murine and human disease

The expression levels of the *VKORC1* transcript were explored in a murine model of MASLD, where animals were treated with an HFD for 30 days. Publicly available data from the Gene Expression Omnibus (GEO, GSE130970) of 78 transcriptomes from patients in differential stages of MASLD were analyzed. The global biobank engine was used to explore the lead MASLD-S single nucleotide polymorphism (rs9934438) in relation to other cardiometabolic traits (Global Biobank Engine, Stanford, CA). GTE_x was used to explore rs9934438 quantitative trait locus.

Statistical analyses

All statistical analyses pertaining MASLD score generation and single-cell RNA-seq were carried out in R 3.5.1, and Plink v.2 was used for genetic analyses. Details are described in detail in the Extended Methods, found in the Supplemental Material, <http://links.lww.com/HEP/I623>.

Ethical approval and consent to participate

Ethical approval was obtained from the Swedish Ethical Review Authority (approval number 10627/18).

Data availability

All sequencing files from HepaRG cell line characterization and Perturb-seq experiments have been deposited on GEO and are publicly available. The repositories are GSE261025, GSM8132774, GSM8132775, GSM8132776, GSM8132777, GSM8132778, GSE238219, GSM7660623, GSM7660624, GSM7660625, GSM7660626, and GSM7660627.

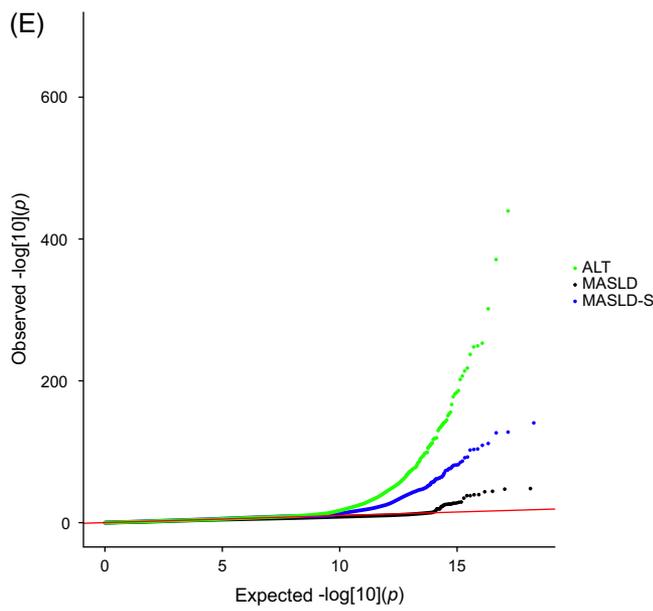
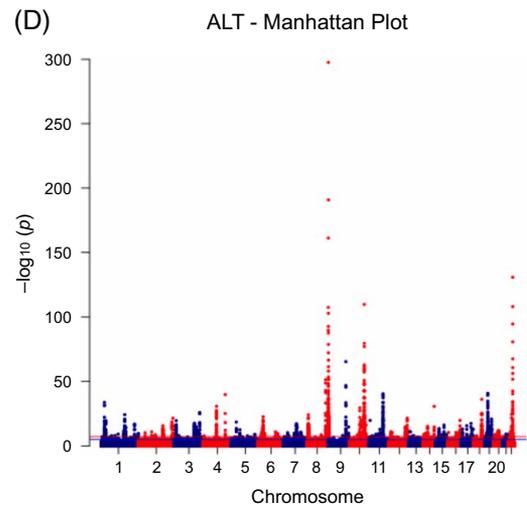
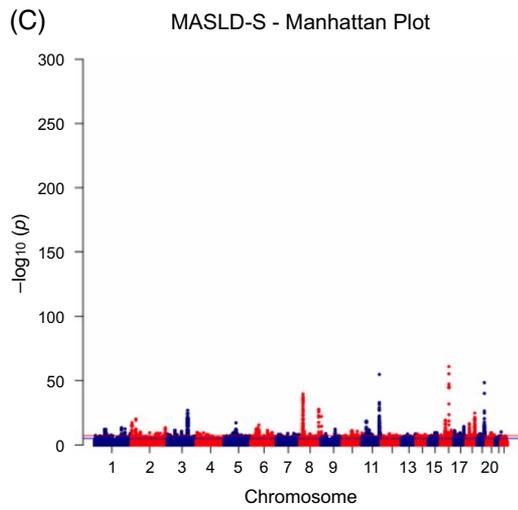
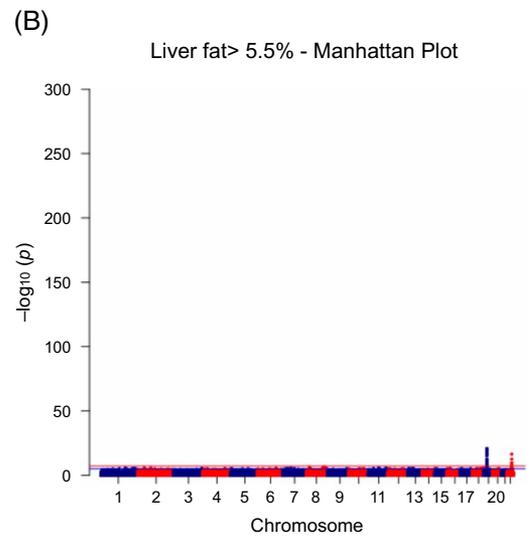
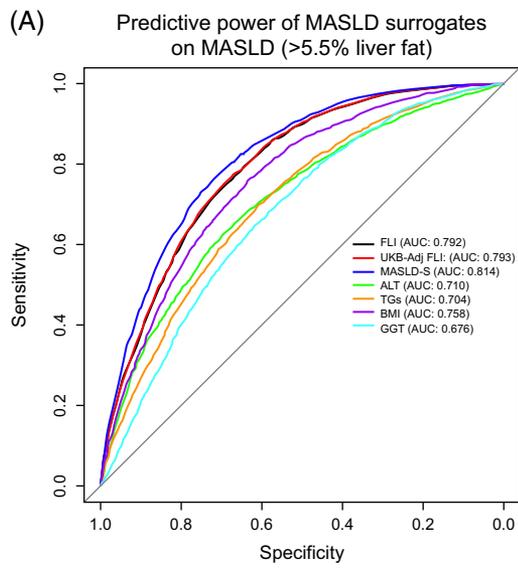


FIGURE 1 Human molecular genetic analyses in the UK Biobank. Nonmoderately or moderately drinking European ancestry British participants were selected for the analyses. (A) receiver operating characteristic curve showing the predictive power of MASLD-S and individual biochemical and anthropometric variables on MASLD status as defined by liver fat > 5.5% in UK biobank. (B) Manhattan plot for the genome-wide association study on MASLD defined as > 5.5% liver fat in UK biobank. (C) Genome-wide association study on MASLD score in UK biobank, visualized using a Manhattan plot. (D) ALT associations from the genome-wide association study in UK biobank, visualized by a Manhattan plot. (E) Q-Q plot for the genome-wide association studies on ALT, MASLD, and MASLD score, plotted together to visualize the differences in significance obtained. y-axes in Manhattan plots are scaled for comparison between the 3 association studies. Abbreviations: BMI, body mass index; FLI, fatty liver index; MASLD, metabolic dysfunction–associated fatty liver disease; MASLD-S, metabolic dysfunction–associated fatty liver disease-score; TG, triglycerides; UKB, UK Biobank.

RESULTS

Anthropometric and biochemical data predict MASLD in UKB

Anthropometric and biochemical variables related to MASLD and cardiometabolic traits were interrogated for their ability to predict MASLD defined as liver fat percentage > 5.5% (2544 MASLD cases and 10,168 controls) using multivariate regression models. The variables that significantly predict MASLD can be found in Table 1. Predictors of MASLD were selected to create a MASLD score using Equation 1.

MASLD score improves MASLD approximation

The power to approximate MASLD using the generated MASLD score (MASLD-S) was assessed using a receiver operating characteristic curve, and the AUC was compared between MASLD-S, FLI, and several

individual anthropometric and biochemical variables. Our results reveal that MASLD-S improves the approximation of MASLD status compared to FLI. Further, MASLD-S outperformed all individual anthropometric and biochemical variables on which the MASLD-S was based (Figure 1A).

The MASLD-S was calculated for a subset of nonmoderate and moderate drinkers in the UKB, and a GWAS was carried out on MASLD-S as a continuous variable. In parallel, GWAS were carried out for liver fat percentage (MRI_UKB) and ALT (qnormALT_UKB) in the same subset of UKB. Our results show numerous associations with liver fat percentage, MASLD-S, and ALT (Figure 1B–D). There is a sizable overlap in loci that are associated with MASLD-S and liver fat percentage. For example, the *PNPLA3* locus is detected in GWAS of MASLD-S, ALT, and liver fat percentage. However, since ALT and MASLD-S use a larger portion of the UKB, there is a substantially larger number of associations for ALT and MASLD-S, compared to liver fat percentage. Another effect of the larger sample size used in the association studies for ALT and

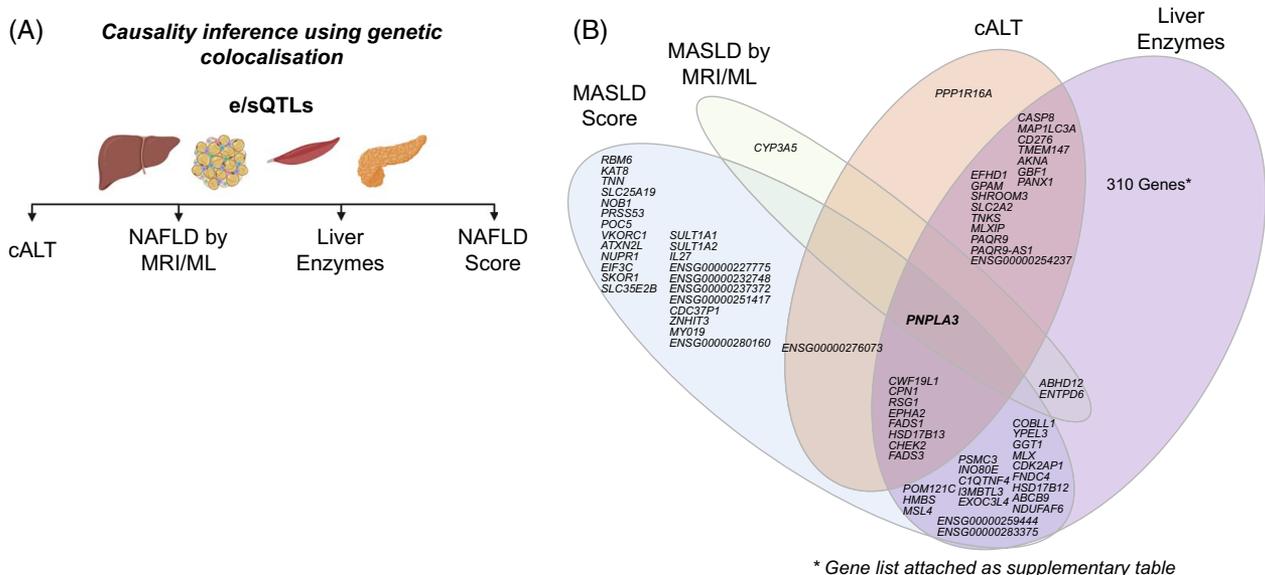


FIGURE 2 Colocalization study of MASLD-S-associated SNPs in the UK Biobank. (A) Strategy for genetic colocalization studies to infer causality of novel putative MASLD genes found from genome-wide association studies with metabolically active tissues in the GTEx (v8) database. Liver enzymes include ALT, ALP, GGT and qnormALT_UKB, MRI/ML MRI_UKB and machine learning MRI, MASLD-S out novel MASLD score, and the MASLD score from Miao and colleagues. (B) Overlap of genes with a significant liver eQTL/sQTL colocalization. GSEA gene set enrichment analysis of colocalized genes can be found in Table 2. Full list of colocalizations can be found in Supplemental Table S5, <http://links.lww.com/HEP/I624>. Abbreviations: cALT, chronic ALT (Alanine Transaminase); ML, machine learning; QTL, quantitative trait locus.

TABLE 2 Results of pathway enrichment analysis for all colocalization signals obtained from MASLD-S associations studies, which demonstrate enrichment in primarily lipid and sterol metabolism pathways

Gene set name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p	FDRq-value
GOBP_LIPID_HOMEOSTASIS [172]	Any process involved in the maintenance of an internal steady state of lipid within an organism or cell. [GOC:BHF, GOC:rl]	8	—	1.22 e ⁻⁷	1.29 e ⁻³
GOMF_OXIDOREDUCTASE_ACTIVITY_ACTING_ON_PAIRED_DONORS_WITH_OXIDATION_OF_A_PAIR_OF_DONORS_RESULTING_IN_THE_REDUCTION_OF_MOLECULAR_OXYGEN_TO_TWO_MOLECULES_OF_WATER [10]	Catalysis of an oxidation-reduction (redox) reaction in which hydrogen or electrons are transferred from each of two donors, and molecular oxygen is reduced to 2 molecules of water. [GOC:mah]	3	—	4.22 e ⁻⁶	2.22 e ⁻²
GOBP_HOMEOSTATIC_PROCESS [1690]	Any biological process involved in the maintenance of an internal steady state. [GOC:jl, ISBN:0395825172]	18	—	1.33 e ⁻⁵	3.29 e ⁻²
GOBP_STEROID_METABOLIC_PROCESS [323]	The chemical reactions and pathways involving steroids, compounds with a 1,2, cyclopentanoperhydrophenanthrene nucleus. [ISBN:0198547684]	8	—	1.35 e ⁻⁵	3.29 e ⁻²
GOBP_LIPID_METABOLIC_PROCESS [1398]	The chemical reactions and pathways involving lipids, compounds soluble in an organic solvent but not, or sparingly, in an aqueous solvent. Includes fatty acids; neutral fats, other fatty-acid esters, and soaps; long-chain (fatty) alcohols and waxes; sphingoids and other long-chain bases; glycolipids, phospholipids and sphingolipids; and carotenes, polyprenols, sterols, terpenes, and other isoprenoids. [GOC:ma]	16	—	1.75 e ⁻⁵	3.29 e ⁻²
GOCC_GAMMA_TUBULIN_COMPLEX [16]	A multiprotein complex composed of gamma-tubulin and other nontubulin proteins. Gamma-tubulin complexes are localized to microtubule organizing centers, and play an important role in the nucleation of microtubules. The number and complexity of non-tubulin proteins associated with these complexes varies between species. [GOC:clt, PMID:12134075]	3	—	1.94 e ⁻⁵	3.29 e ⁻²

GOBP_STEROL_HOMEOSTASIS [100]	Any process involved in the maintenance of an internal steady state of sterol within an organism or cell. [GOC:BHF, GOC:rl]	5	—	2.19 e ⁻⁵	3.29 e ⁻²
GOBP_CHEMICAL_HOMEOSTASIS [1027]	Any biological process involved in the maintenance of an internal steady state of a chemical. [GOC:isa_complete]	13	—	4.07 e ⁻⁵	4.83 e ⁻²
GOBP_CELL_CYCLE [1847]	The progression of biochemical and morphological phases and events that occur in a cell during successive cell replication or nuclear replication events. Canonically, the cell cycle comprises the replication and segregation of genetic material followed by the division of the cell, but in endocycles or syncytial cells nuclear replication or nuclear division may not be followed by cell division. [GOC:go_curators, GOC:mtg_cell_cycle]	18	—	4.27 e ⁻⁵	4.83 e ⁻²
GOMF_STEROL_BINDING [59]	Binding to a sterol, a steroid containing a hydroxy group in the 3 position, closely related to cholestan-3-ol. [GOC:mah]	4	—	4.59 e ⁻⁵	4.38 e ⁻²

TABLE 3 Presentation of the genes targeted in Perturb-seq experiments to create and validate our functional genomic framework

Target genes Perturb-seq
C6orf106
GPAM
LYPLAL1
NCKIPSD
PNPLA3
PPP1R3B
RBM6
TNKS
TRIB1
VKORC1
WDR6

MASLD-S is the typically smaller p -value for these associations, which is visualized by scaled y -axes on Manhattan plots, and the 3 association studies plotted in the same Q-Q plot, Figure 1E. Summary statistics for significant associations can be found in Supplemental Tables S1–S3, <http://links.lww.com/HEP/I624>.

To aid in inferring causality and to prioritize genes for functional follow-up, we assessed GWAS SNPs associated with liver fat percentage and common surrogates through genetic colocalization to eQTL and sQTLs from GTEx (v8) using our custom pipeline.^[12] For this analysis, we used our MASLD-S as well as a previously published score, recently published data on liver enzymes, chronically elevated ALT and MRI/machine learning approaches to approximate MASLD (Figure 2A) to explore the overlap between different approaches to detect genetic associations and causal genes for MASLD.^[8,13–15] Genes demonstrating a significant colocalization to the liver tissue in the GTEx data base were prioritized.

Due to the relative paucity of GWAS data for liver fat percentage, only 4 colocalizations were found for our MRI/machine learning in liver: *PNPLA3* (which is also shared with all MASLD surrogate markers), *CYP3A5*, *ABHD12*, and *ENTPD6*. In contrast, we observed numerous colocalizations originating from GWAS of MASLD surrogates, with sizable overlap between the different surrogates, Figure 2B and Supplemental Table S4, <http://links.lww.com/HEP/I624>. Numerous other genes that have previously been suggested to influence MASLD also show significant colocalization; eg, in or near *GPAM* (ALT), *AKNA* (ALT and MASLD-S), and the *TNKS/PPP1R3B* (ALT). *VKORC1*, a gene previously associated with triglyceride levels and body fat distribution, colocalizes with MASLD-S.^[16,17] We then used colocalized MASLD-S genes as input in a gene set enrichment analysis pathway enrichment analysis, and show that these genes are enriched in processes related to lipid homeostasis, steroid and lipid

metabolism, and sterol homeostasis, Table 2. Colocalized ALT genes, however, are primarily enriched in processes pertaining to organelle organization, small molecule metabolic processes, response to stress, and lipid metabolism, Supplemental Table S5, <http://links.lww.com/HEP/I624>. Importantly, our MASLD-S outperforms ALT in approximating liver fat >5.5%, and variants associated with MASLD-S more often colocalize with lipid metabolism-related eQTLs in the liver ($n=32$ genes) than variants associated with ALT ($n=12$ genes). This provides a rationale for using composite surrogate variables for GWAS of MASLD as these may capture more of the biology of the disease and provide better insight into the natural history of MASLD than single biochemical surrogates.

Collectively, these data not only suggest that creating composite surrogate markers for MASLD may be used to identify putative MASLD genes when there is a paucity of gold standard MRI data, but also that there may be biological differences driving the different associations with surrogate phenotypes, which has implications for the pathogenesis of MASLD.

Establishing a HepaRG cell line suitable for genome editing

To do functional follow-up studies following gene knockdown experiments, we genetically engineered HepaRG cells to stably express dCas9-KRAB, which allows for CRISPRi. The introduction pHR-SFFV-KRAB-dCas9-P2A-mCherry into HepaRG cells allows for transcriptional interference of genes targeted by sgRNAs by KRAB. The resulting cell line (dCas9-KRAB-HepaRG) was used to characterize putative MASLD genes. HepaRG cells underwent single-cell RNA sequencing (scRNA-seq) to characterize the model system and ensure that the introduction of dCas9-KRAB does not alter the function and the ability to differentiate HepaRG cells. dCas9-KRAB-HepaRG cell line was efficiently differentiated using established protocols (Figure 3A), did not differ at the transcriptome level, assessed by scRNA-Seq, regardless of dCas9-KRAB integration, and assumed a hepatocyte-like phenotype on treatment with a differentiation media for 2 weeks (Supplemental Figure S1, <http://links.lww.com/HEP/I625>).

Standard scRNA-seq quality control steps were taken, and revealed that, on differentiation, HepaRG cells increased their expression of mitochondrial genes, while the overall number of genes expressed at detectable levels was marginally decreased, Supplemental Figure S2, <http://links.lww.com/HEP/I626>. The upregulation of mitochondrial genes was not surprising as on differentiation, HepaRG cells have been documented to increase their metabolism while suppressing proliferation. Clustering of single cells showed that there

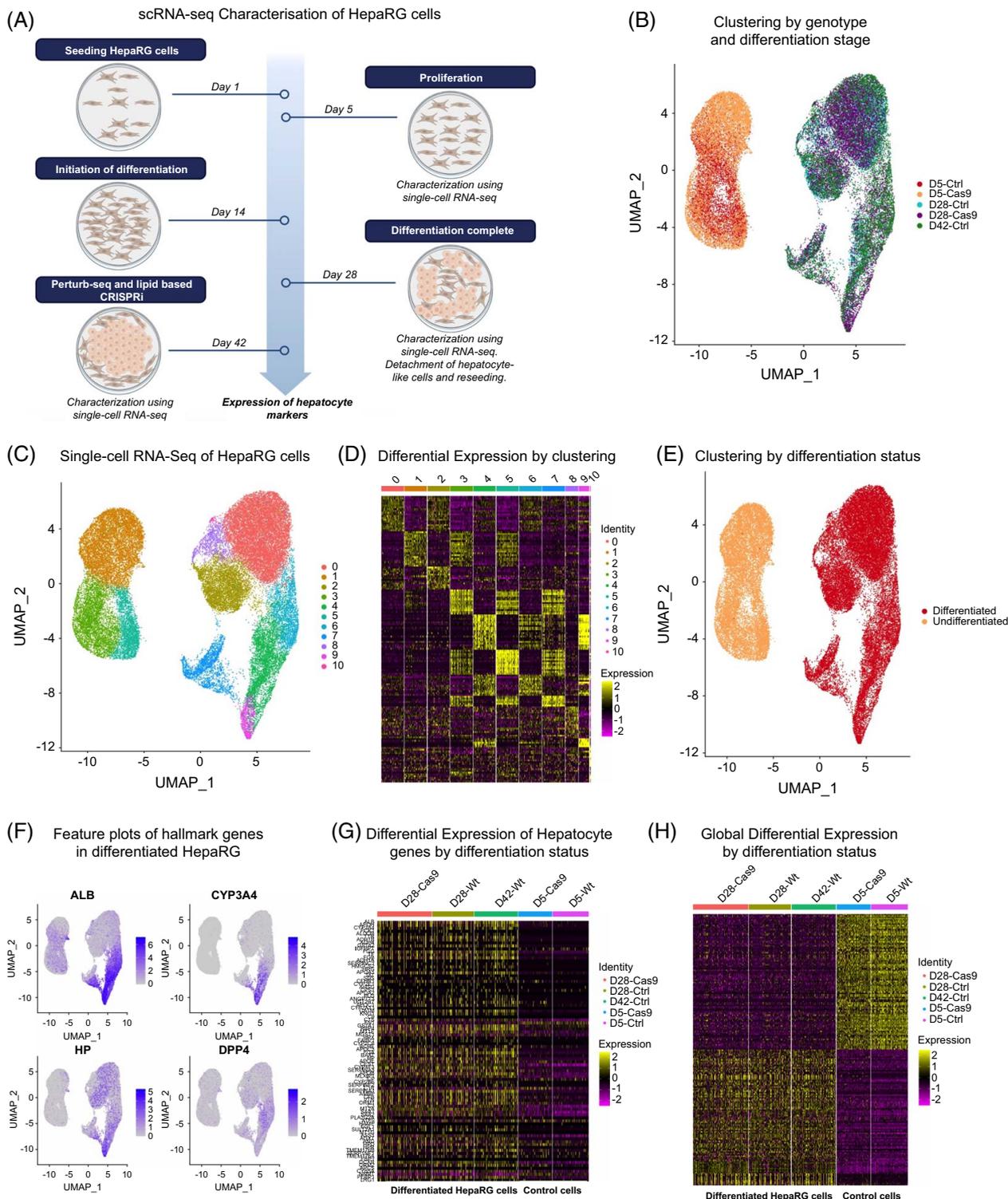


FIGURE 3 Characterization of a HepaRG model system that is genetically engineered to allow for CRISPRi gene-editing. (A) Description of HepaRG culturing, indicating at which point scRNA-seq was used to characterize the model system. (B) Clustering by both genotype and differentiation stage (temporal analysis along the differentiation axis). Data demonstrate that cells efficiently differentiate regardless of genotype (dCas9-KRAB integration) and that cells remain in their differentiated phenotype two weeks after differentiation is complete. This allows for gene editing after complete HepaRG differentiation. (C) Clustering of scRNA-seq data, where proliferative and differentiated cells are plotted together, irrespective of genotype (dCas9-KRAB integration). Data show 11 different clusters divided over two distinct populations of cells. (D) Differential gene expression analyses based on clustering in B. Clusters 1, 3, and 5 belong to undifferentiated cells, whereas the remaining clusters belong to the differentiated HepaRG cells; genes involved in drug-metabolizing pathways, lipid metabolism, hemostasis, and albumin were significantly upregulated in differentiated cells, particularly in clusters 0, 2, 4, 6 and 9. Lists for differentially expressed genes can be found in Supplemental Table S6, <http://links.lww.com/HEP/I624>. (E) Clustering by differentiation status irrespective of genotype. Data demonstrate a perfect clustering of

HepaRG cells by their differentiation status. (F) Expression of hepatocyte hallmark genes *ALB*, *CYP3A5*, *HP-1*, and *DPP4*. Data show an upregulation of these genes on differentiation. (G) Differential expression analyses of genes suggested to define hepatocytes from “the human liver atlas.” Data show that the transcriptional program thought to define hepatocytes is enhanced on differentiation. (H) Global differential expression analyses by differentiation status. Genes upregulated by differentiation are enriched in processes related to small molecule and lipid metabolic processes, mitochondrial processes, and electron transport chain (Supplemental Table S10, <http://links.lww.com/HEP/1624>). Complete lists of differentially expressed genes can be found in Supplemental Table S7, <http://links.lww.com/HEP/1624>. Abbreviations: scRNA, single-cell RNA sequencing; UMAP, uniform manifold approximation and projection.

are no significant differences between wild type (Wt) and genetically engineered dCas9-KRAB-HepaRG cells (Figure 3B), and thus, all cells were analyzed jointly.

Clustering with regard to single-cell transcriptomes of Wt and dCas9-KRAB HepaRG cells revealed 11 distinct clusters; clusters 1, 3, and 5 belong to undifferentiated cells, whereas the remaining clusters belong to differentiated HepaRG cells. Genes involved in the cell cycle, G2M checkpoint, epithelial-mesenchymal transition, and cell division are all more highly expressed in the undifferentiated clusters. In contrast, genes involved in drug-metabolizing pathways, lipid metabolism, hemostasis, and albumin are significantly upregulated in differentiated cells in clusters 0, 2, 4, 6, and 9 (Figure 3C-D, Supplemental Table S6, <http://links.lww.com/HEP/1624>). It is expected that numerous cells undergo apoptosis during the differentiation process. In line with this, cells within clusters 8 and 10 express genes involved in apoptosis, p53, and programmed cell death (Figure 3D), Supplemental Table S6, <http://links.lww.com/HEP/1624>. Cells within cluster 7 seem to consist of a population of cells that may not be fully differentiated, as they highly express some hepatocyte and proliferative markers (Figure 3D and Supplemental Table S6, <http://links.lww.com/HEP/1624>). In summary, HepaRG cells are efficiently differentiated to a hepatocyte-like phenotype as the transcriptome of the cells belonging to clusters 4, 6, 7, and 9 (> 50% of cells) indicate a shift consistent with hepatocyte biology and function.

Differentiated and proliferative HepaRG cells cluster separately, as shown in Figure 3E. We compared the differentiated and undifferentiated cell populations based on a list of hepatocyte markers and the human liver atlas,^[18] regardless of their dCas9-KRAB status. Differentiated cells demonstrate an increased expression of hallmark hepatocyte genes, including *ALB*, *CYP3A4*, *HP*, and *DPP4* (Figure 3F). The expression of a list of hepatocyte genes involved in drug and lipid metabolism was also increased compared to undifferentiated cells (Figure 3G). Next, we analyzed global differential gene expression. Gene set enrichment analysis revealed that differentiated HepaRG cells increase their expression of genes involved in metabolic processes, both in lipid metabolism and the genes within drug metabolism (Figure 3H, Supplemental Figure S3A, B, <http://links.lww.com/HEP/1627>, and

Supplemental Table S7, <http://links.lww.com/HEP/1624>).

CRISPRi screen and Perturb-seq implicate putative causal genes in MASLD

We created a combinatory lipid accumulation-based CRISPRi and Perturb-seq screen in the dCas9-KRAB-HepaRG cell system to investigate putative MASLD genes. We optimized the lipid accumulation-based CRISPRi system by knocking down the lipid droplet-associated protein *PLIN2* to markedly reduce lipid accumulation. dCas9-KRAB-HepaRG cells were transduced with 3 sgRNAs targeting *PLIN2* along with a nontargeting sgRNA as a control, loaded with 400 μ M oleic acid for 24 hours. Next, neutral lipids were stained using Bodipy. Lipid loading was significantly increased after 24 hours of oleic acid treatment, and the efficiency of the sgRNAs was confirmed (Figure 4A-C). mCherry/BFP^{+/+} HepaRG cells were sorted with regard to lipid content after 10 days of gene editing, and genomic DNA was isolated in the most and least lipid-laden cells (the 20th percentile in either tail) (Figure 4D, E). Sequencing of genomic DNA from the most and least of lipid loaded HepaRG cells, as measured by Bodipy staining, revealed a significant enrichment of *PLIN2* sgRNAs in the least lipid-laden cells, indicating that *PLIN2* knock-down indeed impairs lipid accumulation (Figure 4F, G). This experiment served as a proof-of-principle for our CRISPRi screen, which included sgRNAs targeting a small selection of putative MASLD genes, selected based on our human molecular genetic analyses.

Eleven known and putative MASLD genes were selected for tandem CRISPRi and Perturb-seq to explore their role in MASLD development, as measured by HepaRG lipid accumulation and single-cell transcriptional changes. The genes were selected based on (1) their robustness of association to MASLD (amount of evidence if known MASLD gene), (2) emerging evidence for an association without functional validation, and (3) new association with MASLD-S that also demonstrates association to other cardiometabolic traits. The genes can be found in Table 2. Next, sgRNAs directed toward selected MASLD genes were transduced in differentiated dCas9-expressing HepaRG cells for a tandem CRISPRi and Perturb-seq experiment (Figure 5A).

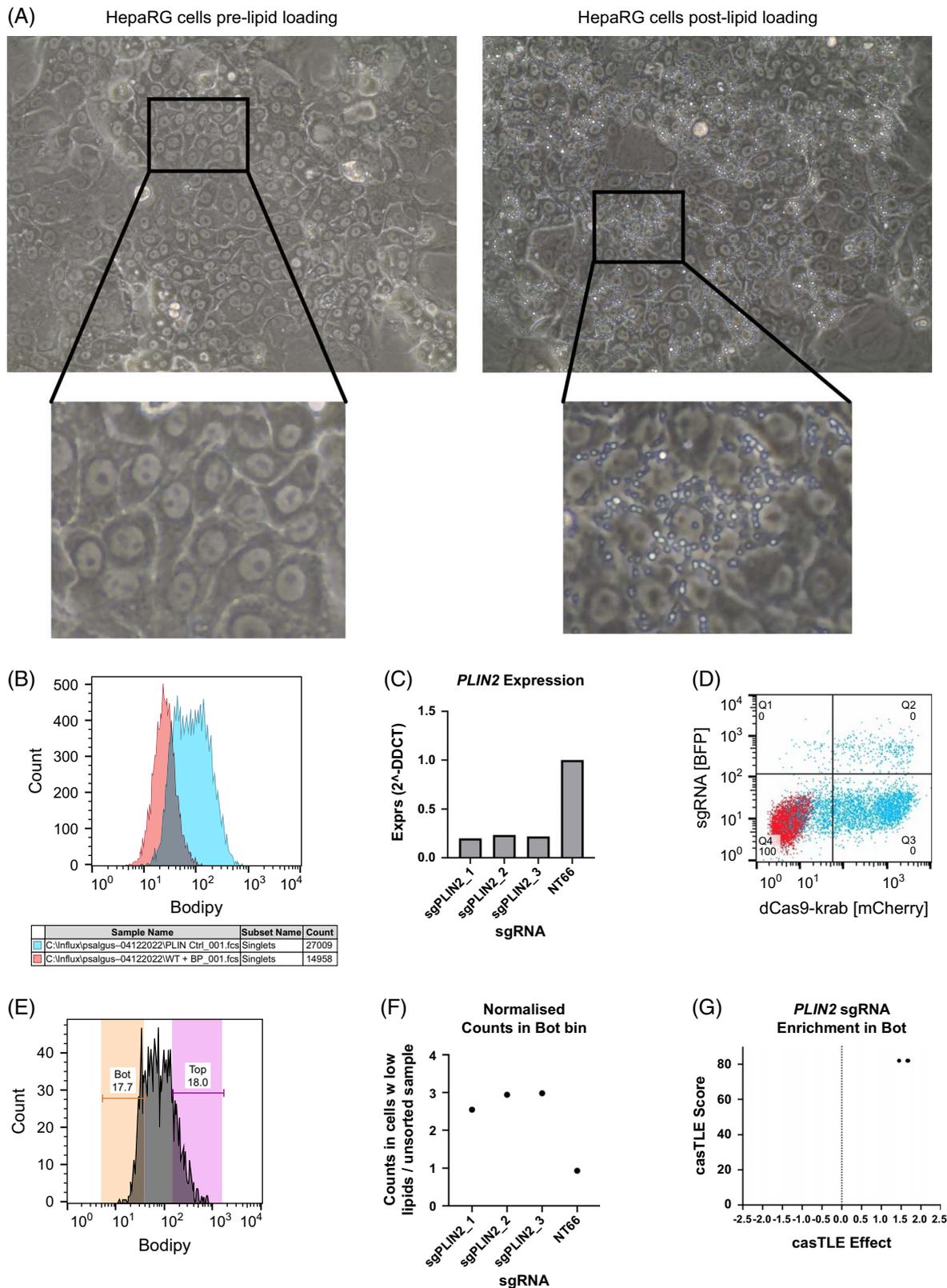


FIGURE 4 Establishment of, and control experiments in a HepaRG cell CRISPRi gene-editing model system with lipid accumulation as readout. (A) Micrographs showing that lipid loading using 400 μM of oleic acid results in significant formation of large lipid droplets. (B) Lipid loaded HepaRG cells were stained with 1 μg/mL Bodipy and analyzed using flow cytometry. Data show that lipid loading (blue histogram) increases the content of neutral lipids within the HepaRG cell compared to non-loaded control cells (red histogram). (C) *PLIN2* was knocked down as a proof-of-principle experiment. *PLIN2* expression was efficiently silenced in our dCas9-KRAB expressing HepaRG cells, and sgRNAs from the v2 Weissman library. (D) Representative gates for sorting gene-edited HepaRG cells (blue), and an untransduced control, negative for both mCherry

and BFP (red). The Q2 gate contains the gene-edited cells, which express dCas9-KRAB and have been efficiently transduced with sgRNAs. (E) Gene-edited HepaRG cells from Q2 were sorted based on their Bodipy content; approximately the top and bottom 18% of cells were sorted, and gDNA was prepared from both extreme populations. gDNA was then sequenced using NGS. (F–G) By assessing the enrichment of *PLIN2* sgRNAs in the cell population with the least intracellular lipids, we find a 2–3 times enrichment of *PLIN2* sgRNAs compared to nontargeting sgRNAs. As one would expect, these data demonstrate hampered lipid accumulation in HepaRG cells that do not express *PLIN2*. The castLE pipeline was also piloted for this purpose, and analyzed data recapitulates simple sgRNA counting and fold change calculations in that we show a 2 times enrichment of *PLIN2* sgRNAs in the least lipid-laden cells compared to what one would expect by chance. Data show that the model system can provide useful information on the effect of genes on lipid accumulation in HepaRG cells, and show that appropriate analysis methods are employed. Abbreviation: sgRNA, small guide RNA.

Cells were loaded with oleic acid and then sorted based on mCherry/BFP^{+/+}, and their lipid accumulation was measured by Bodipy staining. Sequencing of genomic DNA from either extreme population with regard to lipid accumulation (approximately top/bottom 15%) revealed that *VKORC1* and *TNKS* sgRNAs are enriched in the bottom population, whereas *GPAM* and *LYPLAL1* sgRNAs are enriched in the top population. This suggests that *VKORC1* and *TNKS* knockdown reduces lipid accumulation, whereas *GPAM* and *LYPLAL1* knockdown increases lipid accumulation (Figure 5B).

In parallel to the lipid accumulation-based CRISPRi screen, we produced single-cell transcriptomes from all perturbations. The experiment was performed in a total of five 10× genomics single-cell captures, from 2 biological replicates. Single-cell transcriptomes were analyzed using Seurat, and the general quality control data is visualized in Supplemental Figure S3A, <http://links.lww.com/HEP/I627>. While there was no clustering by replicate or sgRNA identity, perturbations produced by the sgRNAs are consistently efficient and specific as only the intended target gene is significantly knocked down (Figure 5C–E). *VKORC1* knockdown produced the most striking transcriptional changes and will be discussed in detail below. *GPAM* knockdown resulted in a downregulation of genes enriched in oxidative phosphorylation and RNA transcription pathways, while the upregulated genes were enriched in pathways pertaining cellular stress, glycolysis, apoptosis, and cell cycle (Supplemental Table S8, <http://links.lww.com/HEP/I624>). *LYPLAL1* knockdown resulted in the downregulation of genes involved in interferon-response, adipogenesis, and oxidative phosphorylation, among others. Genes upregulated by *LYPLAL1* knockdown are enriched in metabolism of heme and blood vessel formation (Supplemental Table S8, <http://links.lww.com/HEP/I624>). We found pathway enrichments in adipogenesis, HDL, and chylomicron metabolism, and estrogen response among genes downregulated on *TNKS* knockdown. Differential gene expression for all perturbations are visualized as heatmaps in Supplemental Figure S4, <http://links.lww.com/HEP/I628>, and Supplemental Table S8, <http://links.lww.com/HEP/I624>.

While we decided to focus on the target gene *VKORC1*—because of its novelty and significant impact on lipid accumulation—we validate one gene (*GPAM*)

influencing lipid accumulation in the opposite direction to *VKORC1* in HepaRG cells. We knocked down *GPAM* using single sgRNA transductions, and recapitulated the findings from the lipid accumulation-based CRISPRi screen, where *GPAM* knockdown results in an increase in lipid accumulation (Supplemental Figure S5A–D, <http://links.lww.com/HEP/I629>).

VKORC1 is involved in the development and progression of hepatosteatosis

Differential gene expression as a result of *VKORC1* knockdown was investigated over the 2 replicates of Perturb-seq experiments using the scMAGeCK package in R. All differentially expressed genes from 2 replicates were investigated for gene set enrichment, and results show that genes enriched in lipid metabolic pathways are downregulated on *VKORC1* knockdown (Figure 5F, G and Supplemental Tables S9–S11, <http://links.lww.com/HEP/I624>). Further, agnostic differential gene expression analyses demonstrate that *VKORC1* knockdown alters the expression of a set of genes related to liver lipid metabolism and insulin resistance (Supplemental Figure S5E, <http://links.lww.com/HEP/I629>). Specifically, under *VKORC1* knockdown conditions, there is a trend for reduced expression in cells of genes involved in lipoprotein production and secretion (*DGAT1*, *DGAT2*, *APOB*, *APOC1*, and *MTTP*) and of the lipid accumulation marker *PLIN2*. Our scRNA-seq data reinforces the notion that *VKORC1* may influence lipid accumulation and *PLIN2* expression since there is a correlation between *PLIN2* and *VKORC1* expression in cells transduced with nontargeting sgRNAs in our Perturb-seq experiments (Supplemental Figure S5F, G, <http://links.lww.com/HEP/I629>).

We construct a protein-protein interaction network using BioGRID to explore what proteins might interact with *VKORC1*. Analyses reveal that there is a physical interaction with apolipoproteins, which reinforces the notion that *VKORC1* may have a previously unexplored role in liver lipid metabolism (Supplemental Figure S6A, <http://links.lww.com/HEP/I630>). We investigate gene set enrichment of all *VKORC1* interactors, and enrichments were found in processes pertaining to lipid homeostasis, oxidoreductase activity, lipid metabolic

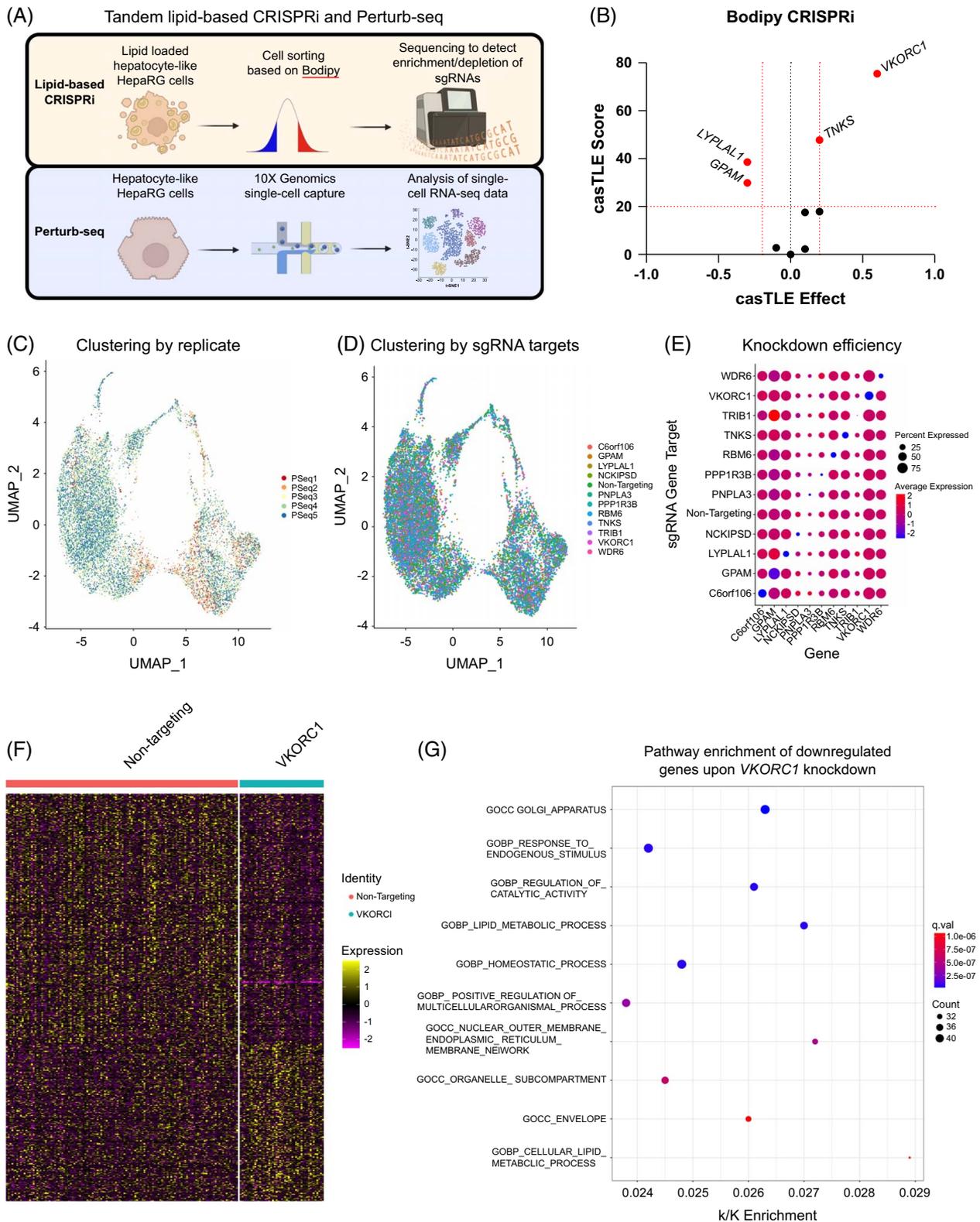


FIGURE 5 Tandem lipid-based CRISPRi and Perturb-seq in HepaRG cells to explore the involvement of genes, suggested by human molecular genetics, in MASLD pathogenesis. (A) Experimental outline of tandem CRISPRi and Perturb-seq in HepaRG cells. HepaRG cells were harvested on day 42 of culturing, as per the protocol described in Figure 3A. (B) Volcano plot, following sequencing of gDNA in the most and least lipid-laden HepaRG cells, where castLE effect and score are plotted against each other. Data demonstrate that the knockdown of *VKORC1* and *TNKS* results in less intracellular lipids. Conversely, the knockdown of genes *GPAM* and *LYPLAL1* increases intracellular lipids. (C, D) Perturb-seq is performed in parallel to our lipid accumulation-based CRISPRi to explore the transcriptomic profiles resulting from a gene knockdown. No major changes in the clustering of gene-edited cells by replicate and sgRNA identity is observed. Data show that replicates are very similar, and sgRNAs have modest effects on the transcriptome that causes the cells to cluster separately. (E) Dotplot is visualizing that the knockdown of sgRNAs

targeting the selected genes is efficient and specific as demonstrated by the blue dots along the diagonal. (F, G) Differential gene expression analyses on *VKORC1* knockdown are carried out using the scMaGeCK R-package, and differentially expressed genes are plotted in a representative heatmap. Results reveal that *VKORC1* knockdown changes the transcriptional landscape, and reduces the gene expression of genes involved in lipid metabolism, Golgi and ER, as well as homeostatic processes. Complete results of differentially expressed genes for all perturbations can be found in Supplemental Figure S4, <http://links.lww.com/HEP/I628>, and Supplemental Table S8, <http://links.lww.com/HEP/I624>. Abbreviations: sgRNA, small guide RNA; UMAP, uniform manifold approximation and projection.

processes, and sterol homeostasis (Supplemental Table S11, <http://links.lww.com/HEP/I624>).

We next performed knockdown experiments of *VKORC1* in HepaRG cells using single sgRNA transductions to confirm our observations from single-cell CRISPRi screens, with a nontargeting sgRNA as control. The knockdown was confirmed using RT-quantitative polymerase chain reaction against *VKORC1* (Figure 6A). We recapitulated the *in vitro* phenotype observed in the single-cell CRISPRi screens, where the reduction of *VKORC1* expression brought about a reduction in *PLIN2* expression, accompanied by a reduction in lipid accumulation as measured by Bodipy using flow cytometry and confocal microscopy (Figure 6B–E).

To better understand the role of *VKORC1* expression in human MASLD, we investigated publicly available data on *VKORC1* transcript levels in a cohort of 78 human livers encompassing the entire spectrum of MASLD. Our analyses suggest a positive association between MASLD activity score, steatosis, and inflammation with *VKORC1* expression (Figure 6F–H). These data suggest that *VKORC1* is involved in the initiation of MASLD; however, *VKORC1* does not seem to be the primary driver of the progression of disease as transcript levels only increase over the lowest grade of disease, and not as grades of disease progress.

We explored the co-expression patterns of *VKORC1* and transcripts of a selection of genes involved in lipid metabolism and fibrosis that are thought to drive disease progression in healthy human liver. Co-expression patterns suggest that *VKORC1* expression correlates with the expression of genes involved in uptake of lipids, as well as in intracellular fatty acid and triglyceride synthesis. Further, *VKORC1* mRNA levels are correlated with transcript levels of collagen and TGF β , which are genes known to promote fibrosis (Supplemental Figure S7A, <http://links.lww.com/HEP/I631>). *VKORC1* was negatively correlated with genes involved in the mobilization of lipids from hepatocytes; *MTTP* and *SREBF1*, suggesting that *VKORC1* expression promotes the intracellular accumulation of lipids in human liver. Collectively, several lines of suggestive data indicate that *VKORC1* may be involved in the initial stages of MASLD natural history.

The *in vitro* MASLD phenotype is also recapitulated in mice fed a high-fat diet for 30 weeks, known to induce MASLD, where both *Vkorc1* and *Plin2* expression is concomitantly increased in animals on high-fat diet (Supplemental Figure S6B, <http://links.lww.com/HEP/I630>).

Further exploration of genetic data and PheWAS revealed a large Lipid droplet-block in the MASLD-S–associated *VKORC1* locus, and that the MASLD-S reducing A allele of lead single nucleotide polymorphism rs9934438 is also associated with reduced lower hip and waist circumference, BMI, and numerous fat mass phenotypes (Supplemental Figure S8A, B, <http://links.lww.com/HEP/I632>). The rs9934438 A allele also shows a protective association with biomarkers of cardiometabolic disease, including lower plasma triglycerides, ApoB, HbA1c, and higher HDL and ApoA (Supplemental Figure S8B, <http://links.lww.com/HEP/I632>). Finally, the rs9934438 A allele is associated with a lower *VKORC1* expression in the liver (GTEx v8 database), reinforcing the observed relationship between *VKORC1* and an *in vitro* MASLD phenotype, as well as the phenotype obtained from *in vivo* models of disease (Supplemental Figure S8C, <http://links.lww.com/HEP/I632>).

In summary, we have demonstrated the usefulness of using MASLD-S as a surrogate marker for MASLD, prioritized candidate MASLD genes from past and present studies using a custom genetic colocalization analysis for functional follow-up. After assigning putative causal genes for functional follow-up coming from GWAS for different MASLD surrogates, we performed a functional CRISPRi screen for lipid accumulation and Perturb-seq transcriptional analysis at a single-cell level, which constitutes a functional genomic framework and allows for interrogation of putative MASLD genes at scale. By using our functional genomics framework, originating from human genetics, moving to functional *in vitro* studies, and later to murine and human disease, we propose that *VKORC1* is implicated in the pathogenesis of MASLD. Our data suggest that *VKORC1* expression is associated with the increase in intracellular accumulation of lipids, and thereby drives the initiation of MASLD development. Investigations can now be expanded to interrogate a large selection of putative causal MASLD genes to further determine the molecular landscape of disease development and progression.

DISCUSSION

In the present study, we generate a MASLD-S that outperforms single variable surrogates when validated against “ground truth” MASLD as defined by >5.5% liver fat obtained from proton density fat fraction from

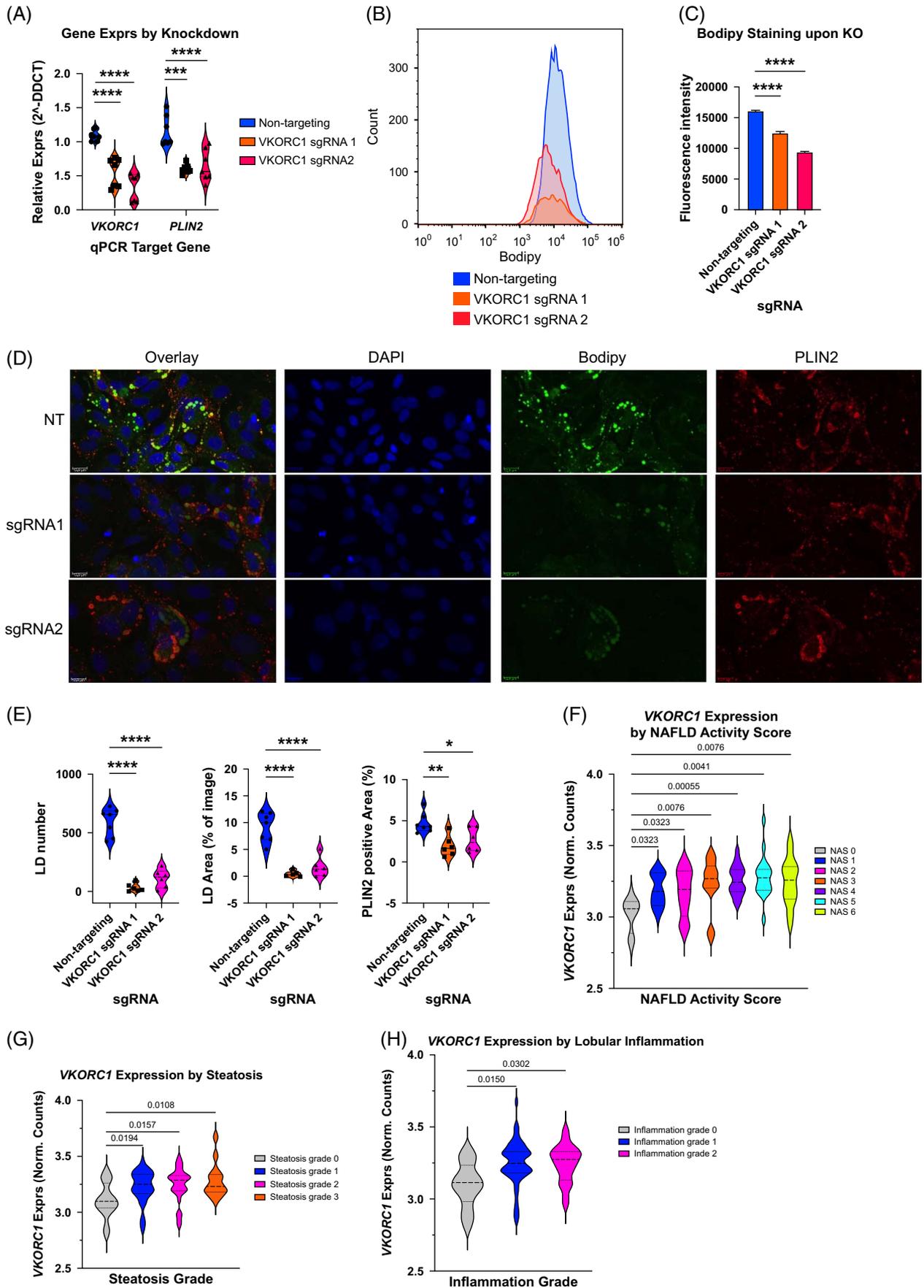


FIGURE 6 Validation experiments of *VKORC1* knockdown in differentiated HepaRG cells and the relationship between *VKORC1* transcript and human disease. (A) Single sgRNA knockdown of *VKORC1* in differentiated HepaRG cells results in a significant knockdown of the *VKORC1* transcript as measured by qPCR. Concomitant with *VKORC1* knockdown, we demonstrate a significant downregulation of the *PLIN2* transcript. (B, C) *VKORC1* knockdown results in the reduction of intracellular neutral lipids by Bodipy staining and flow cytometric analysis. (D, E) Confocal microscopy of HepaRG cells on *VKORC1* knockdown shows a significant reduction in Bodipy neutral lipid staining, lipid droplet number, lipid droplet area, and *PLIN2* positive area. (F–H) By exploring *VKORC1* expression levels in different stages of human disease we demonstrate an upregulation of the *VKORC1* transcript in livers of a higher degree of metabolic dysfunction–associated fatty liver disease activity score, steatosis, and inflammation. N for experimental data is 6–7 replicates, Ordinary one-way ANOVA was performed to compare the nontargeting sgRNA with the *VKORC1* targeting sgRNAs. Total n for human liver samples is 78. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Abbreviations: KD, knockdown; LD, lipid droplet; NT, non-targeting; qPCR, quantitative polymerase chain reaction; sgRNA, small guide RNA.

MRI images in the UKB. By using a new surrogate marker of MASLD for our GWAS, together with colocalization analyses of previous GWAS, we expand the knowledge on the genetic susceptibility to MASLD. We create a functional genomic framework to validate putative MASLD genes and explore the role of a subset of genes on hepatocyte lipid accumulation, single-cell transcriptomes, and murine and human disease.

GWAS have been useful in the identification of common susceptibility variants for various cardiometabolic traits. However, GWAS for MASLD have remained small and underpowered, and therefore, surrogate markers of MASLD have been extensively used, all with their strengths and drawbacks. The FLI does not seem to outperform waist circumference in predicting MASLD,^[4] MASLD may be present without ALT elevation,^[19,20] and high ALT levels could reflect a myriad of liver insults. Thus, FLI and ALT may constitute poor surrogates for MASLD. Our MASLD-S might not only reflect metabolic liver disease but also an insulin resistance phenotype as the score takes into account waist circumference, BMI, HbA1c, and triglyceride levels, which are all also associated with insulin resistance. By integrating anthropometric and biochemical data into a single score, we sought to capture the global etiology of MASLD given its correlation with dyslipidemia, type II diabetes, and obesity. We outperform ALT levels in predicting liver fat in the UKB; however, mitigation of the drawbacks of using ALT measurements as surrogate for MASLD could be achieved by using chronic ALT elevation, which has been described elsewhere.^[14] We find several overlaps in the genetic colocalizations between several MASLD surrogates, which suggests that these surrogates capture a common part of the disease etiology, but also may reflect different aspects of the natural history of MASLD.

We created a HepaRG cell model system suitable for large-scale CRISPRi screening and Perturb-seq to explore putative MASLD genes. We selected a group of both previously associated (*PNPLA3*, *TNKS/PPP1R3B*, *GPAM*, *LYPLAL1*, *TRIB1*) and novel or less established (*WDR6*, *VKORC1*, *RBM6*, *NCKIPSD*, *C6orf106*) MASLD candidate genes to establish a functional genetic framework to interrogate new potential disease genes at scale. We screened the selected genes for their influence on

hepatocyte lipid accumulation in our in vitro gene-editing system, and generated single-cell transcriptomes for these CRISPR-based knockdown perturbations. Our data suggest the involvement of *VKORC1* and *TNKS* (increased lipid content), *LYPLAL1* and *GPAM* (decreased lipid content) on lipid accumulation in hepatocytes, which we further validated for *VKORC1* and *GPAM*. The latter is a well-established MASLD locus whose loss-of-function has been suggested to reduce MASLD, and its knockdown resulted in increased lipid accumulation in HepaRG cells. However, conflicting data do exist with regard to the protective effects of *GPAM* KO.^[21] It is well known that *GPAM* is involved in the TAG synthesis (forming monoacyl glycerol). The effects we see on *GPAM* KD could arise due to the experimental conditions under which these experiments have been conducted. The increased neutral lipid staining may come from the lipid overload in this setting that increases uptake, FA synthesis and reduced lipid excretion. This may be supported by the concomitant reduction in *APOB* and increase in *FASN* mRNA expression (Supplemental Figure S5E, <http://links.lww.com/HEP/I629>). This would facilitate less production of triglyceride-rich lipoprotein particles for excretion, and increased fatty acid synthesis. Further, it is known that *Gpat1* KO results in increased Acyl-CoAs,^[22,23] which we cannot exclude influencing the lipid accumulation (Bodipy) readout. To speculate, the explanation of this phenomenon could be that *GPAM* KD cells compensate for the inability to form new triacylglycerol molecules by excreting less triglyceride-rich lipoproteins, and synthesizing and taking up more free fatty acids. The unexpected results that *GPAM* KD resulted in increased lipid accumulation under lipid loading conditions may be explained by this compensatory mechanism when triacylglycerol synthesis is hampered. In contrast, *VKORC1* knockdown resulted in less lipid accumulation, possibly mediated by lower *PLIN2* levels. The knockdown of *VKORC1* also resulted in the perturbation of the transcriptional landscape of lipid metabolism, and insulin resistance genes like *PNPLA2* (upregulation), and *G6PC*, *PLIN2*, and *INSR* (downregulation) were dysregulated. Finally, we explored the mRNA expression of *VKORC1* in a murine model of disease, where *VKORC1* expression consistently was increased on high-fat diet. We strengthen the notion that low *VKORC1* expression may be protective of human disease development by exploring the expression levels in

relation to the degree of steatosis, MASLD activity score, and inflammation. By using human molecular genetics, we demonstrated that the MASLD-S lowering single nucleotide polymorphism rs9934438 also improves other anthropometric and biochemical cardiometabolic traits, while lowering the expression of the *VKORC1* transcript. Collectively, this suggests a protective role of low *VKORC1* expression in MASLD.

VKORC1 is known to reduce vitamin K to its active form, which promotes the formation of functional clotting factors from proclotting factors. This process is inhibited by warfarin and ultimately results in reduced activation of coagulation factors IX, VII, and prothrombin, which is how warfarin exerts its antithrombotic effects.^[24] Some studies have indeed described an association between thrombotic risk factors and the extent of fibrosis in MASLD.^[25] Similarly, researchers have found elevated and increased activity of coagulation factors in MASLD.^[26] Likewise, it has been observed that there is a higher-than-expected prevalence of MASLD in patients suffering from idiopathic venous thromboembolism.^[27] However, to the best of our knowledge, we provide the first data implicating *VKORC1* in the hepatocyte lipid metabolism, that is also reflected in *VKORC1* expression levels in murine and human disease.

Several genetic variants in the *VKORC1* locus have been described to influence patients' response to warfarin treatment.^[28] However, to the best of our knowledge, no genetic variants in this locus have been described to influence MASLD. By using a composite variable as a proxy for MASLD, we may capture more of the genetic variability contributing to MASLD, than when using single surrogate variables. We may capture more of the metabolic phenotype of MASLD than if only ALT levels had been used since the MASLD-S is made up of liver enzymes, biochemical, and anthropometric variables that are highly correlated with MASLD, obesity, and diabetes. Interestingly, GWAS for BMI,^[29] triglycerides LDL, total cholesterol, and HDL^[16] have identified genetic signals in the *VKORC1* locus, and we find strong colocalization signals for *VKORC1* in the liver (Supplemental Figure S8, <http://links.lww.com/HEP/l632>). Moreover, human PheWAS data show a strong association with BMI, triglycerides, and HDL (among other cardiometabolic values) between variants in *VKORC1*, including a splice donor variant (rs2884737). These data support our transcriptional data that show a dysregulation of lipid metabolism genes on *VKORC1* knockdown, and our protein-protein interaction network suggests a role for *VKORC1* in lipid and cholesterol metabolism.

Collectively, present and previous data provide a potential rationale for the involvement of *VKORC1* in the pathogenesis of MASLD through the regulation of lipid accumulation and cholesterol metabolism in human hepatocytes. To the best of our knowledge, we provide

the first experimental evidence suggesting *VKORC1* as a MASLD susceptibility gene.

In summary, we have expanded our knowledge of the genetic susceptibility for MASLD by using GWA and genetic colocalization studies of surrogate markers of MASLD. Above all, we have established a functional genomic framework to study putative MASLD genes at scale. Large-scale CRISPRi screens have not only paved the way to study genes involved in various cardiometabolic phenotypes,^[30] but also intricate multi-dimensional gene cellular functions. Our efforts have implicated the *VKORC1* gene in the pathogenesis of MASLD. Taken together, this study provides a sound rationale for use of CRISPRi screens to delineate the roles of known and new putative causal risk genes for both MASLD, and other cardiometabolic traits.

Limitations

This work was conceived and executed before the change in guidelines, and nomenclature from NAFLD to MASLD as per the Delphi consensus statement published in December 2023.^[31] Admittedly, adopting a simple 5.5% liver fat from MRI images in nonmoderate and moderate drinkers as a definition of MASLD constitutes a limitation. Adopting a 5.5% liver fat cutoff from MRI images was motivated as a measure of being conservative with regard to MASLD diagnosis. This may, naturally, have influenced these results as the number of MASLD diagnoses is lower than had a 5% cutoff been adopted. The approximation of the MASLD-S is directly influenced by this cutoff and, therefore, also downstream analyses. While the discovery of *VKORC1* as a novel MASLD gene was made in vitro, which can be viewed as a limitation, the strength of this translational framework lies within the replication of *VKORC1* across species, and from human hepatocytes to liver tissue. Finally, the effect of *VKORC1* on adiposity, insulin resistance, and dyslipidemia, which mediates the influence on MASLD, cannot be excluded.

ACKNOWLEDGMENTS

The authors thank all funding bodies contributing to this work, as well as Drs Frank Chenfei Ning and Annelie Falkevall for generously providing liver cDNA samples from mice on high-fat and chow diet.

AUTHOR CONTRIBUTIONS

Peter Saliba-Gustafsson: conceptualization, data acquisition/analysis/interpretation, writing, and funding acquisition. Johanne M. Justesen, Ivan Carcamo-Orive, and Chong Y. Park: conceptualization, data acquisition/analysis/interpretation, and critical reviewing. Amanda Ranta, Hanna M. Björck, Per Eriksson, Anders Franco-Cereceda, and Themistocles L. Assimes: data acquisition/analysis/interpretation and critical reviewing. Disha

Sharma: data analysis/interpretation and critical reviewing. Ewa Bielczyk-Maczynska: data analysis/interpretation and critical reviewing. Jiehan Li, Maider Apodaka, Patricia Aspichueta, Endrina Mujica, Marcel den Hoed, and Thomas Quertermous: data interpretation and critical reviewing. Laeya A. Najmi: data acquisition and critical reviewing. Theresia M. Schurr: data acquisition. Mike Gloudemans: data acquisition. Joshua W. Knowles: conceptualization, data acquisition/analysis/interpretation, critical reviewing, and funding acquisition.

FUNDING INFORMATION

Peter Saliba-Gustafsson is supported by the Swedish Research Council (Vetenskapsrådet), grant number 2018-06580, and the Swedish Heart-Lung Foundation, grant number 20170221. Johanne M. Justesen is funded by grants from the Novo Nordisk Foundation and the Stanford Bio-X Program (NNF17OC0025806). Joshua W. Knowles is supported by the NIH through grants: P30 DK116074 (to the Stanford Diabetes Research Center), R01 DK116750, R01 DK120565, R01 DK106236, and by the American Diabetes Association through grant 1-19-JDF-108. AR is supported by the Finnish Foundation for Cardiovascular Research, Diabetes Research Foundation, Emil Aaltonen Foundation, Ida Montin's Foundation, Biomedicum Helsinki Foundation, Orion Research Foundation, and the Finnish Medical Foundation. The ASAP study was supported by the Swedish Research Council grant number 2020-01442, the Swedish Heart-Lung Foundation grant number 20180451, and a donation by Fredrik Lundberg. Patricia Aspichueta is supported by MCIU/AEI/FEDER, UE (PID2021-124425OB-I00), and Basque Government, Department of Education (IT1476-22). Cell sorting/flow cytometry analysis for this project was done on instruments in the Stanford Shared FACS Facility, funded by NIH S10 Shared Instrument Grant S10RR027431-01. Ivan Carcamo-Orive is supported by a Ikerbasque Research Fellowship, funded by the EU (H2020-MSCA-COFUND-2020-101034228-WOLF-RAM2), and grant PID2023-148986OB-I00 funded by MCIU/AEI/FEDER-EU.

CONFLICTS OF INTEREST

Johanne M. Justesen is employed by Novo Nordisk. Joshua W. Knowles consults for Arrowhead and Mammoth. The remaining authors have no conflicts to report.

ORCID

Peter Saliba-Gustafsson <https://orcid.org/0000-0002-7807-9009>

Johanne M. Justesen <https://orcid.org/0000-0002-0484-8522>

Amanda Ranta <https://orcid.org/0000-0002-2612-5165>

Disha Sharma <https://orcid.org/0000-0001-9486-2709>

Ewa Bielczyk-Maczynska <https://orcid.org/0000-0002-0558-1188>

Maider Apodaka <https://orcid.org/0000-0002-8459-4752>

Patricia Aspichueta <https://orcid.org/0000-0002-3553-1755>

Hanna M. Björck <https://orcid.org/0000-0002-9155-3609>

Per Eriksson <https://orcid.org/0000-0002-5635-2692>

Theresia M. Schurr <https://orcid.org/0000-0002-6573-4959>

Anders Franco-Cereceda <https://orcid.org/0000-0002-3427-9455>

Mike Gloudemans <https://orcid.org/0000-0002-9924-9943>

Endrina Mujica <https://orcid.org/0000-0003-1348-8765>

Marcel den Hoed <https://orcid.org/0000-0001-8081-428X>

Themistocles L. Assimes <https://orcid.org/0000-0003-2349-0009>

Thomas Quertermous <https://orcid.org/0000-0002-7645-9067>

Ivan Carcamo-Orive <https://orcid.org/0000-0001-8823-4925>

Chong Y. Park <https://orcid.org/0009-0004-9401-2535>

Joshua W. Knowles <https://orcid.org/0000-0003-1922-7240>

REFERENCES

1. Shang Y, Nasr P, Widman L, Hagström H. Risk of cardiovascular disease and loss in life expectancy in NAFLD. *Hepatology*. 2022; 76:1495–505.
2. Duell PB, Welty FK, Miller M, Chait A, Hammond G, Ahmad Z, et al. Nonalcoholic fatty liver disease and cardiovascular risk: A scientific statement from the American Heart Association. *Arterioscler Thromb Vasc Biol*. 2022;42:e168–85.
3. Mancina RM, Sasidharan K, Lindblom A, Wei Y, Ciociola E, Jamialahmadi O, et al. PSD3 downregulation confers protection against fatty liver disease. *Nat Metab*. 2022;4:60–75.
4. Sveinbjornsson G, Ulfarsson MO, Thorolfsdottir RB, Jonsson BA, Einarsson E, Gunnlaugsson G, et al. Multiomics study of nonalcoholic fatty liver disease. *Nat Genet*. 2022;54: 1652–63.
5. Chen Y, Du X, Kuppa A, Feitosa MF, Bielak LF, O'Connell JR, et al. Genome-wide association meta-analysis identifies 17 loci associated with nonalcoholic fatty liver disease. *Nat Genet*. 2023; 55:1640–50.
6. Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol*. 2006;6:33.
7. Motamed N, Sohrabi M, Ajdarkosh H, Hemmasi G, Maadi M, Sayeedian FS, et al. Fatty liver index vs waist circumference for predicting non-alcoholic fatty liver disease. *World J Gastroenterol*. 2016;22:3023–30.

8. Haas ME, Pirruccello JP, Friedman SN, Wang M, Emdin CA, Ajmera VH, et al. Machine learning enables new insights into genetic contributions to liver fat accumulation. *Cell Genom.* 2021; 1:100066.
9. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods.* 2017;14:297–301.
10. Replogle JM, Norman TM, Xu A, Hussmann JA, Chen J, Cogan JZ, et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat Biotechnol.* 2020;38:954–61.
11. Morgens DW, Deans RM, Li A, Bassik MC. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat Biotechnol.* 2016;34:634–6.
12. Gloudemans MJ, Balliu B, Nachun D, Durrant MG, Ingelsson E, Wabitsch M, et al. Integration of genetic colocalizations with physiological and pharmacological perturbations identifies cardiometabolic disease genes [Internet]. 2021. Accessed February 20, 2022;2021.09.28.21264208. <https://www.medrxiv.org/content/10.1101/2021.09.28.21264208v1>.
13. Pazoki R, Vujkovic M, Elliott J, Evangelou E, Gill D, Ghanbari M, et al. Genetic analysis in European ancestry individuals identifies 517 loci associated with liver enzymes. *Nat Commun.* 2021;12:2579.
14. Vujkovic M, Ramdas S, Lorenz KM, Guo X, Darlay R, Cordell HJ, et al. A multiancestry genome-wide association study of unexplained chronic ALT elevation as a proxy for nonalcoholic fatty liver disease with histological and radiological validation. *Nat Genet.* 2022;54:761–1.
15. Miao Z, Garske KM, Pan DZ, Koka A, Kaminska D, Männistö V, et al. Identification of 90 NAFLD GWAS loci and establishment of NAFLD PRS and causal role of NAFLD in coronary artery disease. *HGG Adv.* 2022;3:100056.
16. Graham SE, Clarke SL, Wu K-HH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature.* 2021;600:675–9.
17. Rask-Andersen M, Karlsson T, Ek WE, Johansson Å. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nat Commun.* 2019;10:339.
18. Aizarani N, Saviano A, Sagar, Maily L, Durand S, Herman JS, et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature.* 2019;572:199–204.
19. Matteoni CA, Younossi ZM, Gramlich T, Boparai N, Liu YC, McCullough AJ. Nonalcoholic fatty liver disease: A spectrum of clinical and pathological severity. *Gastroenterology.* 1999;116: 1413–9.
20. Bacon BR, Farahvash MJ, Janney CG, Neuschwander-Tetri BA. Nonalcoholic steatohepatitis: An expanded clinical entity. *Gastroenterology.* 1994;107:1103–9.
21. Yazdi M, Ahnmark A, William-Olsson L, Snaith M, Turner N, Osla F, et al. The role of mitochondrial glycerol-3-phosphate acyltransferase-1 in regulating lipid and glucose homeostasis in high-fat diet fed mice. *Biochem Biophys Res Commun.* 2008; 369:1065–70.
22. Hammond LE, Neschen S, Romanelli AJ, Cline GW, Ilkayeva OR, Shulman GI, et al. Mitochondrial glycerol-3-phosphate acyltransferase-1 is essential in liver for the metabolism of excess acyl-CoAs. *J Biol Chem.* 2005;280:25629–36.
23. Neschen S, Morino K, Hammond LE, Zhang D, Liu Z-X, Romanelli AJ, et al. Prevention of hepatic steatosis and hepatic insulin resistance in mitochondrial acyl-CoA:glycerol-sn-3-phosphate acyltransferase 1 knockout mice. *Cell Metab.* 2005;2: 55–65.
24. Li T, Chang C-Y, Jin D-Y, Lin P-J, Khvorova A, Stafford DW. Identification of the gene for vitamin K epoxide reductase. *Nature.* 2004;427:541–4.
25. Assy N, Bekirov I, Mejritsky Y, Solomon L, Szvalb S, Hussein O. Association between thrombotic risk factors and extent of fibrosis in patients with non-alcoholic fatty liver diseases. *World J Gastroenterol.* 2005;11:5834–9.
26. Kotronen A, Joutsu-Korhonen L, Sevastianova K, Bergholm R, Hakkarainen A, Pietiläinen KH, et al. Increased coagulation factor VIII, IX, XI and XII activities in non-alcoholic fatty liver disease. *Liver International.* 2011;31:176–83.
27. Di Minno MND, Tufano A, Rusolillo A, Di Minno G, Tarantino G. High prevalence of nonalcoholic fatty liver in patients with idiopathic venous thromboembolism. *World J Gastroenterol.* 2010;16:6119–22.
28. Rost S, Fregin A, Ivaskevicius V, Conzelmann E, Hörtnagel K, Pelz H-J, et al. Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. *Nature.* 2004; 427:537–41.
29. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet.* 2018;27:3641–9.
30. Bielczyk-Maczynska E, Sharma D, Blencowe M, Saliba-Gustafsson P, Gloudemans MJ, Yang X, et al. CROP-Seq: A single-cell CRISPRi platform for characterizing candidate genes relevant to metabolic disorders in human adipocytes. *Am J Physiol Cell Physiol.* 2023;325:C648–60.
31. Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *Hepatology.* 2023;78: 1966–86.

How to cite this article: Saliba-Gustafsson P, Justesen JM, Ranta A, Sharma D, Bielczyk-Maczynska E, Li J, et al. A functional genomic framework to elucidate novel causal metabolic dysfunction-associated fatty liver disease genes. *Hepatology.* 2025;82:165–183. <https://doi.org/10.1097/HEP.0000000000001066>