A growing number of long noncoding RNAs (lncRNAs) have emerged as vital metabolic regulators. However, most human lncRNAs are nonconserved and highly tissue specific, vastly limiting our ability to identify human lncRNA metabolic regulators (hLMRs). In this study, we established a pipeline to identify putative hLMRs that are metabolically sensitive, disease relevant, and population applicable. We first progressively processed multilevel human transcriptome data to select liver lncRNAs that exhibit highly dynamic expression in the general population, show differential expression in a nonalcoholic fatty liver disease (NAFLD) population, and respond to dietary intervention in a small NAFLD cohort. We then experimentally demonstrated the responsiveness of selected hepatic lncRNAs to defined metabolic milieus in a liver-specific humanized mouse model. Furthermore, by extracting a concise list of protein-coding genes that are persistently correlated with lncRNAs in general and NAFLD populations, we predicted the specific function for each hLMR. Using gain- and loss-of-function approaches in humanized mice as well as ectopic expression in conventional mice, we validated the regulatory role of one nonconserved hLMR in cholesterol metabolism by coordinating with an RNA-binding protein, PTBP1, to modulate the transcription of cholesterol synthesis genes. Our work overcame the heterogeneity intrinsic to human data to enable the efficient identification and functional definition of disease-relevant human lncRNAs in metabolic homeostasis.
Identification of human long noncoding RNAs associated with nonalcoholic fatty liver disease and metabolic homeostasis

Xiangbo Ruan,1 Ping Li,1 Yonghe Ma,1 Cheng-fei Jiang,1 Yi Chen,1 Yu Shi,1 Nikhil Gupta,1 Fayaz Seifuddin,2 Mehdi Pirooznia,2 Yasuyuki Ohnishi,3 Nao Yoneda,3 Megumi Nishiwaki,3,4 Gabrijela Dumbovic,5 John L. Rinn,5 Yuichiro Higuchi,1 Kenji Kawai,6 Hiroshi Suemizu,3 and Haiming Cao1

1Cardiovascular Branch, National Heart, Lung, and Blood Institute, NIH, Bethesda, Maryland, USA. 2Bioinformatics and Computational Biology, National Heart, Lung, and Blood Institute, NIH, Bethesda, Maryland, USA. 3Laboratory Animal Research Department, Biomedical Research Laboratory, Central Institute for Experimental Animals, Kawasaki, Japan. 4Technical Service Department, CLEA Japan Inc., Shizuoka, Japan. 5Department of Biochemistry and BioFrontiers, University of Colorado Boulder, Boulder, Colorado, USA. 6Department Pathology Analysis Center, Central Institute for Experimental Animals, Kawasaki, Japan.

A growing number of long noncoding RNAs (IncRNAs) have emerged as vital metabolic regulators. However, most human IncRNAs are nonconserved and highly tissue specific, vastly limiting our ability to identify human IncRNA metabolic regulators (hLMRs). In this study, we established a pipeline to identify putative hLMRs that are metabolically sensitive, disease relevant, and population applicable. We first progressively processed multilevel human transcriptome data to select liver IncRNAs that exhibit highly dynamic expression in the general population, show differential expression in a nonalcoholic fatty liver disease (NAFLD) population, and respond to dietary intervention in a small NAFLD cohort. We then experimentally demonstrated the responsiveness of selected hepatic IncRNAs to defined metabolic milieu in a liver-specific humanized mouse model. Furthermore, by extracting a concise list of protein-coding genes that are persistently correlated with IncRNAs in general and NAFLD populations, we predicted the specific function for each hLMR. Using gain- and loss-of-function approaches in humanized mice as well as ectopic expression in conventional mice, we validated the regulatory role of one nonconserved hLMR in cholesterol metabolism by coordinating with an RNA-binding protein, PTBP1, to modulate the transcription of cholesterol synthesis genes. Our work overcame the heterogeneity intrinsic to human data to enable the efficient identification and functional definition of disease-relevant human IncRNAs in metabolic homeostasis.

Introduction

In the past 2 decades, metabolic diseases have reached epidemic proportions globally (1, 2), underscoring the urgent need for a better understanding of metabolism regulation in humans. Although significant progress has been made in mapping the makeup and wiring diagram of metabolic pathways in recent years, substantial gaps remain in our understanding of the underlying pathophysiology of major metabolic diseases, such as obesity, nonalcoholic fatty liver disease (NAFLD), and type 2 diabetes. Given that maintenance of metabolic homeostasis requires high-level coordination at the cellular, organ, and organismal levels (3–5), such a level of complexity often cannot be adequately modeled by cultured cells and needs to be studied in vivo systems, such as mice, which have the advantage of being conducive to genetic manipulation. Although these animal-based experimental studies are essential to capture mechanistic and physiologically relevant understanding of the role of metabolism in health and disease, there are limitations due to species differences and imperfect disease models. These species-distinct differences are further underscored by our incomplete understanding of the human genome, particularly the enormous landscape of noncoding regions. Because 2% of the human genome is sufficient to encode all protein-coding genes, the vast majority of the genome is noncoding and was once considered to be made of gene deserts. It is now well-established that most of the noncoding regions can be transcribed, giving rise to approximately 60,000 long noncoding RNAs (IncRNAs) (6), which would equate to 3 times the number of protein-coding genes. Growing evidence supports the notion that IncRNAs play a regulatory role in systemic energy metabolism in mice. For example, we have shown that a liver-enriched IncRNA, IncLSTR, regulates systemic lipid metabolism (7), and a second IncRNA, IncLGR, regulates glycogen content in mice (8). Robust mouse IncRNA metabolic regulators (mLMRs), such as Lexis, Mexis, and Blnc1 (9–11), have also been reported by many groups, and this list continues to expand. Furthermore, several genome-wide transcriptome analyses in mice have identified hundreds of potential mLMRs in key metabolic organs, suggesting that mouse IncRNAs constitute an additional dimension of metabolic regulation (12, 13). If human IncRNAs exercise a similar function, studying their metabolic function could help systemically uncover novel regulatory mechanisms of human metabolism and expand our understanding of how metabolic disease is initiated and progress.
Despite their enormous potential, it is currently extremely difficult to define human lncRNA metabolic regulators (hLMRs), in part because of the multitude of challenges in assigning functions to lncRNAs in general, and in determining the metabolic function of human lncRNAs in particular. Current knowledge and technology limit our ability to identify and characterize lncRNA functions, especially relative to the progress that has been made for protein-coding genes. Since our current understanding of the sequence-function relationship of lncRNAs is very poor, we cannot use sequence features, such as functional domains in protein-coding genes, to place lncRNAs in a biological context (14).

To address this challenge in the context of energy metabolism, considerable efforts have been devoted to identifying mLMRs by analyzing the regulatory information of a lncRNA in response to various conditions to inform its function. For example, we have developed a pipeline to identify mLMRs based on their regulations by multiple pathophysiologically representative metabolic conditions in mice (12). Sallam et al. identified 2 mLMRs of cholesterol metabolism based on their regulation by liver X receptor (LXR), a well-established transcription factor in cholesterol homeostasis (10, 11). The valuable information yielded by these extensive studies could in theory help identify hLMRs if lncRNAs were as conserved as mRNAs. Surprisingly and intriguingly, however, over 80% of human lncRNAs are not conserved (15, 16), and most human lncRNAs cannot be found in mice and vice versa. Thus, most human lncRNAs belong to a unique class of molecules whose regulatory information has to be directly derived from human studies. Although clinical RNA-Seq data of human metabolic tissues are emerging and accumulating, their numbers and the metabolic conditions under which they are collected are limited. More importantly, unlike data generated in inbred mouse strains under well-controlled experimental conditions, the gene expression levels in humans are significantly affected by genetic heterogeneity (17) and environmental factors. With so many complicating factors involved, it is evidently not a trivial task to retrieve hLMR signals from human data that truly reflect metabolic responses, or are metabolically sensitive, without losing their significance to the general population, and an approach that is specifically suitable for human lncRNAs is needed. Furthermore, for protein-coding genes, the definitive functional validation is routinely carried out in research animals, particularly in mice, by creating gain- or loss-of-function models. Given that most human lncRNAs are nonconserved, their physiological function cannot be directly studied in conventional mice, and an in vivo model is needed to experimentally characterize putative hLMRs in a physiologically relevant setting.

In order to leverage the accumulating clinical studies to understand the pathophysiological importance of lncRNAs in human metabolism, we have established an effective strategy to retrieve a list of broadly representative and metabolically sensitive human lncRNAs from human transcriptome data. We further refined our selection based on the regulation of these human lncRNAs by defined metabolic conditions in a humanized mouse model, and most importantly, experimentally defined the in vivo role of a nonconserved hLMR in cholesterol metabolism in the humanized mouse model.

### Results

#### Identification of hLMRs

To take advantage of the currently available human data while overcoming its limitations, we combined a variety of human studies and a humanized mouse model to establish a practical platform to identify population-applicable, metabolically sensitive, and disease-relevant hLMRs (Figure 1). In order to maintain the broad representation of selected IncRNAs in the general population, we started our analysis with human liver RNA-Seq data from the Genotype-Tissue Expression (GTEx) project (18, 19). Although GTEx RNA-Seq data harbor valuable information for distinct tissue-relevant genes, the data lack deep clinical phenotyping for identifying differentially expressed genes linked to metabolic disease or therapy. Interestingly, information on gene expression variability or dynamics in a specific cell type or tissue has been utilized to infer potential roles in pathophysiology and diseases (20–22). We postulate that lncRNAs with high expression variability in key metabolic tissues such as the liver in the general population from GTEx could be potentially metabolically sensitive and disease relevant. Therefore, we determined the gene expression variability in the livers of individuals in GTEx. To extend the coverage for human lncRNAs, we have recently established an updated and comprehensive lncRNA database, lncRNA Knowledgebase (lncRNAKB) (23), which we used to map human lncRNAs throughout this study (see Methods). In total, the coefficients of variation for 16,906 genes, including 2665 lncRNAs expressed in the liver samples (Figure 2A), were calculated and ranked in 4 quartiles. We then assigned the protein-coding genes in the top and bottom quartile separately to the category of complex diseases using the DAVID gene functional annotation tool. As shown in Figure 2B, there was a significant enrichment of multiple disease categories in the top quartile of protein-coding genes from the liver, whereas very few were found in the bottom quartile. This result suggests that hepatic genes with high dynamic expression in the general human population might be conditionally responsive and susceptible to a variety of complex disease conditions, particularly cardiometabolic diseases (Figure 2B). Thus, the 943 lncRNAs included in the top quartile (Supplemental Table 1; supplemental material available online with this article; https://doi.org/10.1172/JCI136336DS1) could be metabolically sensitive and potentially function as human lncRNA regulators in the development of cardiometabolic diseases (Figure 1). In order to leverage the accumulating clinical studies to understand the pathophysiological importance of lncRNAs in human metabolism, we have established an effective strategy to retrieve a list of broadly representative and metabolically sensitive human lncRNAs from human transcriptome data. We further refined our selection based on the regulation of these human lncRNAs by defined metabolic conditions in a humanized mouse model, and most importantly, experimentally defined the in vivo role of a nonconserved hLMR in cholesterol metabolism in the humanized mouse model.

Although the 348 lncRNAs have potential implications for both general and disease populations, their observed changes in expression levels could be affected by genetic heterogeneity inherent to population-based human studies. To minimize such effects and further enrich lncRNAs whose expression levels are indeed...
metabolically sensitive, we next analyzed a liver RNA-Seq data set generated from liver biopsies of 7 obese NAFLD patients before and after a short-term intervention with a low-carbohydrate diet (25). A low-carbohydrate diet is known to cause rapid and robust reductions of liver fat as well as changes in the expression of genes in multiple metabolic pathways (25). Out of the 348 IncRNAs, 77 were also regulated by the low-carbohydrate dietary intervention in this small cohort (Supplemental Table 3 and Supplemental Table 4). Thus, by integrating gene expression dynamics in the general population and gene regulation in observational and interventional human studies related to NAFLD, we effectively identified a group of population-applicable, metabolically sensitive, and disease-relevant human IncRNAs, which we refer to as putative hLMRs (Figure 1, identification steps 1–3).

In recent years, it has become a common practice to predict a lncRNA’s function based on its correlated protein-coding genes in a large transcriptome data set (12, 26–28). On many occasions, however, the number of such correlated genes could be very large, and the predicted functions could be very broad. Additionally, the correlated genes often vary significantly in distinct populations. To overcome these limitations, we extracted genes persistently correlated with a lncRNA in multiple independent data sets, which we postulated could remove some protein-coding genes that were spuriously coregulated with a lncRNA and enrich the lncRNA’s specifically correlated genes. Specifically, we intersected protein-coding genes correlated with a lncRNA in the general GTEx population and samples from the NAFLD study described above. As a result, we identified a concise list of correlated genes of each hLMR in a metabolic disease–relevant setting. The Gene Ontology (GO) analysis using these lists indicated that our identified hLMRs may function in diverse metabolic pathways, such as fatty acid oxidation, the cholesterol biosynthetic process, and glucose metabolism (Figure 3B and Supplemental Table 4). 

**Metabolic regulation of hLMRs in humanized mice.** After filtering out confounding variables to identify metabolically sensitive hLMRs and predict their functions, we next asked whether we could further investigate their metabolic responses under a physiologically relevant and well-controlled experimental condition. Indeed, we recently found that a liver-specific humanized mouse model, which was produced by human hepatocytes from a single donor and kept in a defined environment in an animal facility (29), is suitable for studying the regulation of human-specific IncRNAs (30). We thus performed RNA-Seq analysis to identify differentially expressed human genes in the liver-specific humanized mice subjected to a fasting-refeeding regime, which involves the 2 extreme ends of caloric cycles and is known to regulate nearly all key metabolic genes in vivo (Supplemental Table 5). As expected, the expression of protein-coding genes involved in fatty acid oxidation and gluconeogenesis was upregulated by fasting and downregulated by refeeding, whereas genes in the lipogenesis pathway showed the opposite pattern (Figure 4A and Supplemental Figure 1A), supporting the proper response of human genes to nutrient and hormone levels in the humanized liver. Furthermore, we noticed that a significant portion of differentially expressed human genes during fasting and refeeding overlapped with those in the NAFLD and the low-carbohydrate dietary intervention analysis (Figure 4B). These results further support the notion that the humanized liver maintains a proper gene expression response to metabolic milieu as the human liver does. Finally, we found that 20 out of 77 of the liver hLMRs that we identified were regulated by feeding cycles in the humanized mice (Figure 4C). Indeed, the specific regulations of these 20 hLMRs were largely in line with their predicted function. For example, lnckb.38556, which is downregulated by fasting and recovered by refeeding, is predicted to function in biosynthesis of cholesterol.

Taken together, by performing stepwise selections of lncRNAs from multiple data sets representing the general population, disease populations, interventional studies, and well-controlled experiments in humanized mice, we established a list of potential hLMRs that are population applicable, metabolically sensitive, and disease relevant (Figure 1, identification steps 1–4). Furthermore, by extracting protein-coding genes that persistently correlated with IncRNAs in independent populations, we were able to generate a concise list of genes that could be utilized to infer the function of
each hepatic hLMR for downstream analysis (Figure 1, prediction steps).

**Regulation of hepatic cholesterol biosynthetic pathway by hLMR1 in the humanized liver.** Among the 20 liver hLMRs, we used lnckb.38556, which we refer to as hLMR1, as an example to experimentally validate our selection and prediction process. hLMR1 is annotated as a 5-exon intergenic lncRNA located in chromosome 3 of the human genome with the full-length transcript around 2 kb, and bioinformatic prediction supported its poor coding potential (Supplemental Figure 1, B and C). No homolog of hLMR1 could be identified in mice by a BLAST search, suggesting it is a nonconserved human lncRNA. To further characterize the gene structure and transcript isoforms of hLMR1, we took advantage of the PacBio long-read RNA-Seq data (31) and found several hLMR1 isoforms with the same or very similar 5’ end and exactly the same 3’ end compared with the reference (Supplemental Figure 1D). Our cloned full-length hLMR1 included most exons detected in the PacBio isoforms, and an in vitro translation assay confirmed that this transcript was noncoding (Supplemental Figure 1, D and E). To avoid omission of functional isoforms of hLMR1, we designed our real-time PCR (qPCR) primers and shRNA targeting hLMR1 within its last exon, which was shared by all isoforms (Supplemental Figure 1D).

Our qPCR results using human tissue cDNA panels showed that hLMR1 was exclusively expressed in human liver tissue (Supplemental Figure 1F) and had a conservatively estimated copy number per cell of 56. Subcellular fractionation analysis using humanized liver tissues found that hLMR1 was distributed in the cytoplasm and nucleus, with more hLMR1 in the nuclear fraction (Supplemental Figure 2A). Single-molecule RNA (smRNA) FISH using cultured primary human hepatocytes further confirmed the predominantly nuclear localization of hLMR1 (Supplemental Figure 2B). RNA-Seq analysis of humanized livers showed that hLMR1 was downregulated during fasting and recovered upon refeeding (Figure 4C), which
The idea that hLMR1 positively regulates the cholesterol biosynthetic pathway, as predicted. We noticed that depletion of hLMR1 had no effect on the expression of \textit{PAQR9}, the close neighbor gene of hLMR1 (Figure 5A), indicating it is unlikely that hLMR1 functions through modulating the expression of \textit{PAQR9}.

Given that our pipeline was designed to identify hLMRs implicated in the general population, we next tested whether the regulatory effects of hLMR1 could be observed in a different genetic background. A knockdown experiment as described above was hence performed in humanized mice prepared with hepatocytes from a second independent and ethnically different donor. As shown in Figure 5B, with a similar knockdown efficiency of hLMR1 in these mice, we found significant downregulation of \textit{SC5D}, \textit{LSS}, \textit{FDPS}, and \textit{HMGCS1}, which was consistent with the result we observed in mice produced with the first donor. Taken together, our results indicate that hLMR1 is critical to maintain the expression of cholesterol biosynthetic genes in human populations.

To further study the regulatory effects of hLMR1, we next asked whether overexpression of hLMR1 could promote the expression of the 6 crucial genes in the cholesterol biosynthetic pathway whose expressions correlated with hLMR1 in our GO term analysis, 4 genes (\textit{SC5D}, \textit{FDPS}, \textit{LSS}, and \textit{HMGCS1}) showed decreased expression by more than 50% upon depletion of hLMR1 in the humanized livers (Figure 5A). This result thus supported the idea that hLMR1 positively regulates the cholesterol biosynthetic pathway, as predicted. We noticed that depletion of hLMR1 had no effect on the expression of \textit{PAQR9}, the close neighbor gene of hLMR1 (Figure 5A), indicating it is unlikely that hLMR1 functions through modulating the expression of \textit{PAQR9}.

Given that our pipeline was designed to identify hLMRs implicated in the general population, we next tested whether the regulatory effects of hLMR1 could be observed in a different genetic background. A knockdown experiment as described above was hence performed in humanized mice prepared with hepatocytes from a second independent and ethnically different donor. As shown in Figure 5B, with a similar knockdown efficiency of hLMR1 in these mice, we found significant downregulation of \textit{SC5D}, \textit{LSS}, \textit{FDPS}, and \textit{HMGCS1}, which was consistent with the result we observed in mice produced with the first donor. Taken together, our results indicate that hLMR1 is critical to maintain the expression of cholesterol biosynthetic genes in human populations.

To further study the regulatory effects of hLMR1, we next asked whether overexpression of hLMR1 could promote the expression of...
The regulatory effects of hLMR1 on cholesterol biosynthetic genes encouraged us to examine whether hLMR1 could modulate cholesterol levels in the humanized mice. The humanized livers in these mice are chimeric, and it is technically challenging to ascertain the specific impact of human hepatocytes. We thus used an immunoaffinity approach to specifically isolate human ApoB-containing lipoproteins from the plasma of humanized mice (Supplemental Figure 2E and Methods). Using this method, we found that depletion of hLMR1 in humanized livers led to a 40% decrease in human LDL and VLDL cholesterol levels compared with control humanized mice (Figure 5D). Taken together, our bioinformatic analyses using large-scale human data and functional analyses in humanized livers suggest that a nonconserved human lncRNA, hLMR1, is a crucial regulator of the hepatic cholesterol biosynthetic pathway and could play a critical role in the maintenance of cholesterol homeostasis in humans.

hLMR1 coordinates PTBP1 to promote the transcription of cholesterol biosynthetic genes. To explore the molecular mechanism mediating the regulatory effects of hLMR1, we first performed RNA poly II ChIP analysis to determine the transcriptional activities of hLMR1 target genes in the humanized livers. As shown in Figure 6A, humanized livers with depletion of hLMR1 showed significantly lower enrichment of RNA poly II on the transcription start site of human SC5D, FDPS, LSS, and HMGCS1. These data suggest that hLMR1 regulates the expression of cholesterol biosynthetic genes by promoting their transcription, in line with the relative enrichment of hLMR1 in the nuclear fraction (Supplemental Figure 2, A and B). To further explore how hLMR1 regulates the transcription of its target genes, we next performed RNA pulldown combined with mass spectrometry analysis to identify proteins that interact with hLMR1. This strategy successfully identified PTBPI, an RNA-binding protein that regulates almost all steps of cholesterol biosynthetic pathway. To this end, we took advantage of the CRISPR activation (CRISPRa) tool to enhance the expression of endogenous hLMR1 specifically in human hepatocytes of the humanized liver. As shown in Figure 5C, adenovirus-mediated expression of CRISPRa targeting hLMR1 in humanized mice induced the expression level of hLMR1 by 4-fold, whereas the expression of PAQR9 was not affected. Corroborating the result of our knockdown experiments, we found that induction of hLMR1 by CRISPRa resulted in significant upregulation of SC5D, FDPS, LSS, and HMGCS1. This data further suggests that hLMR1 is not only necessary for maintaining the expression of cholesterol biosynthetic genes, but also sufficient to promote their expression.
Figure 5. Regulation of cholesterol metabolism by hLMR1 in humanized mice. (A) Gene expression in humanized mice receiving adenovirus for control (sh-lacZ, n = 5) or knockdown of hLMR1 (sh-hLMR1, n = 5). (B) Gene expression in humanized mice (second donor) receiving adenovirus for control (sh-lacZ, n = 4) or knockdown of hLMR1 (sh-hLMR1, n = 4). (C) Gene expression in humanized mice receiving adenovirus for CRISPRa-control (n = 4) or CRISPRa-hLMR1 (n = 3). (D) Relative cholesterol levels in human ApoB-containing lipoproteins purified from the plasma of humanized mice receiving adenovirus for control (sh-lacZ, n = 7) or knockdown of hLMR1 (sh-hLMR1, n = 10). Data are shown as the mean ± SEM, *P < 0.05 by 2-tailed, unpaired Student’s t test.
Figure 6. hLMR1 coordinates PTBP1 to promote the transcription of cholesterol biosynthetic genes. (A) RNA polymer II ChIP analyses in liver tissues of humanized mice receiving adenovirus for control (sh-lacZ, *n* = 3) or knockdown of hLMR1 (sh-hLMR1, *n* = 3). (B) Left: Western blot analysis of PTBP1 in hLMR1 pulldown, right: gene expression in PTBP1 RIP (RNA IP) in humanized liver. (C) HMGCS1 promoter–driven luciferase reporter assay in 293A cells (*n* = 3 for each group). Data are representative results of 3 independent experiments and are shown as the mean ± SEM, *P* < 0.05 by 1-way ANOVA with post hoc Tukey’s honestly significant difference test. (D) PTBP1 ChIP analyses in liver tissues of humanized mice receiving adenovirus for control (sh-lacZ, *n* = 3) or knockdown of hLMR1 (sh-hLMR1, *n* = 3). Data in A and D are shown as the mean ± SEM, *P* < 0.05 by 2-tailed, unpaired Student’s *t* test.
suggests that mouse Ptbp1
Sc5d
expression of hLMR1 resulted in marked upregulation of
ble to that in human hepatocytes. As shown in Figure 7B, we found
target genes. As shown in Figure 7C. Furthermore, when Ptbp1 was depleted, ectopic expression of hLMR1 was no longer able to induce the expression of its target genes, suggesting that the regulatory effects of hLMR1 were dependent on the expression of Ptbp1 (Supplemental Figure 4B).
Finally, both plasma and hepatic cholesterol levels were increased upon expression of hLMR1 (Supplemental Figure 4C). These data, in combination with our results in humanized mice, support the crucial role of hLMR1-Ptbp1 complex in cholesterol metabolism.

Hepatic expression of hLMR1 is associated with cholesterol levels in human population. To further explore the impact of hLMR1 on cholesterol metabolism in the human population, we used expression quantitative trait loci (eQTL) and GWAS integrative analysis to determine the association between hepatic expression of hLMR1 and lipid levels in the general population (33). As shown in Figure 8, encouragingly, we found that several cis eQTLs of hLMR1 overlapped with GWAS loci for total cholesterol levels. A summary-data-based Mendelian randomization (SMR) analysis, which tests whether the effect size of an SNP on the phenotype is mediated by gene expression using data from GWAS and eQTL studies (34), was then performed to further determine whether the overlapped eQTL/GWAS loci were functionally related. The analysis passed both SMR and heterogeneity in dependent instruments (HEIDI) tests (SMR \( P < 0.05 \) and HEIDI \( P > 0.05 \), see Methods), suggesting the hepatic expression level of hLMR1 might contribute to the regulation of cholesterol levels in humans.

Discussion
Human IncRNAs constitute a significant portion of the human transcriptome and have been shown to play a critical role in diverse biological processes. Despite this, the role of human IncRNAs in systemic energy metabolism is poorly understood, in part because of the challenging nature of identifying hLMRs and defining their metabolic function in a physiologically relevant context. In this work, we established an integrated bioinformatic and experimental pipeline to identify hLMRs based on their regulatory information in the general population, patients with metabolic disease, and a humanized mouse model (Figure 1). We also adapted an improved approach to infer the function of hLMRs. Finally, we provided a proof-of-principle example that the metabolic function of human IncRNAs could be successfully defined in a humanized mouse model, where we validated that a non-conserved human IncRNA exhibited its predicted function in cholesterol metabolism and confirmed it as a legitimate hLMR.

Compared to data generated in inbred mice under well-defined conditions, human gene expression data are intrinsically very noisy because of multiple confounding factors, particularly the diverse genetic background and environmental factors. On the other hand, steps to remove the impact of genetic and other confounding factors have the risk of restricting the significance of the findings to a specific subpopulation. To address these challenges, our work has established a platform to identify putative
hLMRs that are metabolically sensitive, disease relevant, and population applicable. Several steps in our selection procedure were designed to better retrieve specific signals against the background noises in human data and at the same time retain the significance of identified IncRNAs to the general population. For example, the responsiveness to multiple disease or metabolic conditions could enrich IncRNAs that are metabolically sensitive, and regulations in small cohorts and the general population alike could ensure that the functional importance of selected IncRNAs has broad implications.

Our approach is entirely scalable and can readily integrate any new RNA-Seq data of human metabolic organs under any additional metabolic conditions. These data are expected to become available en masse with the widespread use of RNA-Seq technology. Although our selective procedure could potentially enrich functional IncRNAs that play a role in metabolism, it does not immediately confirm functionality. Thus, this work also underscores the fundamental importance of a humanized mouse model for the definitive validation of the physiological importance of any putative hLMRs, which are mostly non-conserved. Expression of human IncRNAs in conventional mice, as we have utilized in the work, is a useful alternative to humanized mice, and it may mitigate some of the limitations.
for how PTBP1 exercises its function in cells and animals. PTBP1 is a well-established regulator of splicing, but its identified functions have expanded rapidly in recent years. A number of lncRNAs have been recently reported to interact with PTBP1 to regulate its known functions or expand its functional space. For example, lncRNA LUCAT1 interacts with PTBP1 to facilitate alternative splicing of a set of DNA damage–related genes (37), and lncRNAs MEG3 and UCA1 have been shown to form a complex with PTBP1 to regulate RNA decay (38) or enhance RNA stability (39), respectively. Several lncRNAs have also been recently reported to bind to PTBP1 to regulate transcription (40, 41), which represents a relatively newly identified function for PTBP1.

PTBP1 is universally and abundantly expressed in most tissues, but loss-of-function studies in animals have showed that it often exerts a tissue-specific function (42, 43), raising the question of how the tissue-specific regulation of PTBP1 function is achieved. Compared to protein-coding genes, the expression of lncRNAs is much more cell type– and tissue-specific, and their specific interactions with PTBP1 might confer tissue specificity to PTBP1 function. Consistent with this notion, hLMR1 is highly specific to the liver, the primary site for cholesterol synthesis in the body. With the growing number of human lncRNAs shown to bind to PTBP1, studying the in vivo function of these lncRNAs is expected to enhance our understanding of the functional importance of PTBP1 in physiology and human diseases.

LncRNAs are the largest and probably also the least conserved transcript class in the human genome (15). With the growing importance of human lncRNAs in biology and physiology, defining hLMRs could drive profound changes in the way we study energy metabolism experimentally and understand metabolic disease conceptually, and we hope that efforts to define a functional hLMR could help chart the way forward.

Methods

Bioinformatics analysis

Analysis pipeline for human RNA-Seq data. FASTQ read files were cleaned using TrimGalore. Reads were then aligned using HISAT2 to an index created using the GRCh38 genome and the lncRNAKB annotation (23). Aligned.sam files were then compressed into .bam files and sorted using Sambamba sort. featureCounts from the subread package was used to count reads/fragments aligned to genes (at the exon feature level). Samples with less than 1,000,000 reads aligned were removed. Individual count files were merged into a single count file for each data set using a python script. In each data set, a threshold of more than 1 cpm in at least 50% of the samples was applied to identify differentially expressed lncRNAs.
samples was applied before further analysis. The RNA-Seq data sets for cross-sectional human studies were retrieved from BioProject PRJNA512027 for the 139 liver samples from the NAFLD population. The RNA-Seq data sets for interventional human studies were retrieved from BioProject PRJNA420975 for the 7 paired liver samples of the low-carbohydrate dietary intervention in NAFLD in the NCIB's Sequence Read Archive (SRA) database.

Analysis pipeline for humanized mice RNA-Seq data. RNA-Seq data have been deposited in NCBI's Gene Expression Omnibus (GEO, GSE130525). A combined human and mouse annotation was created by combining the IncRNAKb gene transfer format (GTF) annotation with the RefSeq GCF000001635.26 GTF annotation. A combined human and mouse genome was created by combining the GRCh38.p12 genome sequence; the primary assembly was obtained from the Genencode website and the GRcm38.p6 genome sequence obtained from the RefSeq FTP site. RNA-Seq cleaning, alignment, sorting, quantification, and filtering were conducted in the same way as human RNA-Seq analysis using this index and genome file. Human gene counts were extracted from the resulting count file by removing reads mapped to mouse contigs. A more than 1 cpm in 50% of samples cutoff was applied separately to human gene counts and mouse gene counts.

Differential expression analysis and principal component analysis. A combined raw count file generated by the subread feature-Counts tool for each data set was imported into R. Technical replicates were combined using the collapseReplicates function from the DESeq2 package. The variance stabilizing transform from the DESeq2 package was applied to the count data from before and after combining technical replicates (if there were technical replicates) before conducting principal component analysis. The top 2 PCs were graphed to visualize clustering between experimental groups. DESeq2 was used with non-normalized count data to find differentially expressed genes between experimental groups. Covariates were controlled for by adding them to experimental design if available. A cutoff of logFC more than 0.4 and \( q \) less than 0.05 was used for differential expression for all the liver samples.

Gene variability analysis. The gene expression profiles of the human liver from GTEx version 7 includes samples from Caucasian, Asian, and African ancestry (primarily Caucasian). An expression cutoff of more than 1 cpm in 50% samples was applied in each tissue to reduce mapped genes to 16,906 expressed genes, including 2665 IncRNA genes in the liver. For each of the expressed genes, we quantified expression variability by calculating its coefficient of variation \( \eta \) across all the available samples in each tissue. These were subsequently ranked and split into quartiles. Diseases category analysis was performed by using the DAVID gene functional annotation tool.

Gene correlation analysis. Count data for each data set was variance stabilizing transform (VST) normalized using the DESeq2 R package. After VST normalization, 10 hidden technical factors were calculated per data set using the probabilistic estimation of expression residuals (PEER) software package and were used in linear regression as covariates to correct VST normalized expression data. The mean expression of each gene was added back to the expression residual. The correlation between IncRNAs and protein-coding genes was analyzed by Pearson's method using the normalized data in the GTEx human population and metabolic disease–relevant population. Significantly correlated protein-coding genes with \( P \) less than 0.05 in the liver were used for further GO analysis.

Mice with a humanized liver

TK-NOG mice, in which a herpes simplex virus type 1 thymidine kinase (TK) transgene under a mouse albumin promoter is expressed within the liver of highly immune-deficient NOG mice, were obtained from Taconic Biosciences. The TK converts an antiviral medication ganciclovir (GCV) into a toxic product that allows selective elimination of TK-positive cells in vivo. The cryopreserved primary human hepatocytes were obtained from Lonza. The humanized TK-NOG mice were prepared as previously described (29). Briefly, the TK-NOG mice at 8–9 weeks old received an i.p. injection of GCV at a dose of 25 mg/kg. One week later, 50 \( \mu \)L volume of 1 × 10^6 human primary hepatocytes suspended in HBSS solution were transplanted via intrasplenic injection. The serum human albumin in the mice was measured as an index of the extent of human hepatocyte replacement 8-12 weeks after transplantation. Humanized TK-NOG mice with serum human albumin levels above 0.5 mg/mL were used for experiments in which human hepatic genes could be reliably detected by qPCR. For the fasting-refeeding study, humanized mice were produced, and the experiment was carried out at Central Institute for Experimental Animals (CIEA). Humanized mice for the rest of the study were produced and analyzed at National Heart, Lung, and Blood Institute (NHLBI). For the fasting-refeeding study, humanized TK-NOG mice were allowed free access to food (fed, harvested at around 9–10 am) or subjected to a 24-hour food withdrawal (fasting, from -9:00 am to 9:00 am), or subjected to a 24-hour food withdrawal followed by a 4-hour refeeding (refeeding from -9:00 am to 1:00 pm) before tissue harvest. Animal data were excluded from experiments based on preestablished criteria of visible abnormal liver structure during sample harvest or other health issues such as fighting wounds or infections. According to the variability of metabolic parameters, group size was determined based on previous studies using similar assays within the laboratory and pilot experiments. Experimenters were not blinded to treatment group.

RNA extraction, RNA-Seq, qPCR analysis

Total RNA was isolated from liver tissues using TRIzol reagent (Invitrogen). After Turbo DNA-free DNase treatment (Ambion), the construction of strand-specific sequencing libraries using Illumina TruSeq RNA sample Prep kit and the sequencing were performed at the NHLBI DNA Sequencing and Genomics Core. The reverse transcription was carried out with SuperScript III First-Strand Synthesis system (Invitrogen) using 1 \( \mu \)g of RNA. Quantitative real-time RT-PCR was performed on a ViiA 7 Real-Time PCR System (Applied Biosystems Inc.) The PCR program was 2 minutes 30 seconds at 95°C for enzyme activation, 40 cycles of 15 seconds at 95°C, and 1 minute at 60°C. Melting curve analysis was performed to confirm the real-time PCR products. For detecting the expressions of human genes in humanized liver samples, human-specific primers were designed and quantitation was normalized to human 16S rRNA levels. For detecting the expressions of mouse genes in regular mice (C57BL/6), 18S rRNA was used as the internal control. For detecting the expression of hLMR1 in human tissues, the Human MTC Panel I (Takara Bio Inc., 636742) was used. The full primer sequences used are provided in Supplemental Figure 5.

In vitro translation

The in vitro translation analysis was performed using the TnT Quick Coupled Transcription/Translation System from Promega.
by following the manufacturer’s protocol. The translated protein was visualized by using IRDye streptavidin (LI-COR), which detected the biotinylated lysine incorporated into the translated proteins. The open reading frame of yellow fluorescent protein (YFP) was used as a positive control.

smRNA FISH

smRNA FISH and microscopy were performed according to published protocols and as described previously (44, 45). Probe set against hLMR1 labeled with Quasar 570 was custom designed using LGC Biosearch Technologies’ Stellaris online probe designer (version 4.2). GAPDH was hybridized with a predesigned probe set labeled with Quasar 670 (LGC Biosearch Technologies).

Primary human hepatocytes (Lonza) were seeded on collagen-coated glass coverslips (Thermo Fisher Scientific, NC0636242) in full growth media. The next day, coverslips were washed 2 times with PBS and fixed in 3.7% formaldehyde (Sigma-Aldrich) in PBS for 10 minutes at room temperature (RT). After fixation, cells were permeabilized in 70% ethanol at 4°C for 1 hour. Coverslips were incubated in prehybridization buffer (10% deionized formamide [Agilent] in wash buffer A [LGC Biosearch Technologies]) for 5 minutes at RT. Cells were hybridized with 50 μL of hybridization buffer (LGC Biosearch Technologies) supplemented with 10% deionized formamide (1:100 dilution of smRNA FISH probes) overnight at 37°C in a humid chamber. The next day, cells were washed with wash buffer (wash buffer A with 10% deionized formamide) for 30 minutes at 37°C, followed by another wash (wash buffer A with 10% deionized formamide) containing Hoechst DNA stain (1:1000; Thermo Fisher Scientific) for 30 minutes at 37°C. Coverslips were washed with wash buffer B (LGC Biosearch Technologies) for 5 minutes at RT, equilibrated for 5 minutes in base glucose buffer (2× SSC, 0.4% glucose solution, 20 mM Tris pH 8.0 in RNase-free H2O), and then 5 minutes in base glucose buffer supplemented with 1:100 dilution of glucose oxidase (stock 3.7 mg/mL) and catalase (stock 4 mg/mL). The coverslips were mounted with ProlongGlass (Invitrogen) on a glass slide and left to cure overnight.

Z-stacks (250 nm z-step) capturing the entire cell volume were acquired with a GE wide-field DeltaVision Elite microscope with an Olympus UPlanSApo 100×/1.40 numerical aperture oil objective lens and a PICO Edge sCMOS camera using appropriate filters. The 3-dimensional stacks were deconvolved with the built-in DeltaVision SoftWoRx Imaging software. Fiji was used to generate maximum intensity projections and for signal quantification.

Adenovirus production and in vivo adenovirus administration

The shRNAs for hLMR1 and mouse Ptbp1 were designed using the following sequences (hLMR1 shRNA: CCTTCACAGCTCTGCCTAA; mouse Ptbp1 shRNA: GACCCGTCGCTGAGATCAT). The hairpin template oligonucleotides were synthesized by Integrated DNA Technologies and were subsequently cloned into the adenovirus vector of the pAD/Block-it system (Invitrogen) according to the manufacturer’s protocols. The overexpression construct of hLMR1 was generated by PCR amplifying the sequence of ENST00000476385.1 using a human liver cDNA sample. The sequence was subsequently cloned into pAd5 vector for virus packaging. Adenoviruses were amplified in HEK293A cells and purified by CsCl gradient centrifugation. Purified viruses were desalted with PD10 columns (GE Healthcare Life Sciences) and titered with Adeno-X Rapid Titer Kit (Clontech). Adenoviruses were delivered into mice intravenously at 5 × 108 PFU/mouse for overexpression experiments or at 1 × 108 PFU/mouse for knockdown experiments. After 7 days, tissue samples were harvested for further analysis after mice had been fasted for 24 hours (~9:00 am-9:00 am) and followed by a 4-hour refeeding (~9:00 am to 1:00 pm).

CRISPRa in humanized mice

CRISPRa assay was conducted using the SAM system (http://sam.genome-engineering.org/protocols/). The sgRNA for hLMR1 was designed using the following sequence: forward, CACCGGACAGACAGAGACACT; reverse, AAACAGTCGCCTCCGTCC. Briefly, the template was ligated into the sgRNA (MS2) cloning backbone (Addgene, 61424) using Golden-Gate reaction (SAM) after Golden-Gate annealing. The expression cassettes of dCas9-VP64 (Addgene, 61422) and MS2-P65-HSF1 (Addgene, 61423) were subcloned into pAd5 vector (Invitrogen) for virus packaging. SgRNA elements with the U6 promoter were amplified and subsequently cloned into pAd/PL adenovirus vector (Invitrogen) for virus packaging. Viruses were amplified, desalted, and titered as described above. Three adenoviruses (1:1:1) were delivered into humanized mice intravenously at a total of 5 × 10⁸ PFU/mouse. After 7 days, tissue samples were harvested for further analysis after mice were under food withdrawal for around 5 hours (from ~9:00 am to 2:00 pm).

RNA pulldown assay

RNA pulldown was performed as described previously (46). Briefly, biotin-labeled RNAs were transcribed in vitro using the Biotin RNA Labeling Mix and T7 RNA polymerase (Ambion) and purified with the RNeasy Mini Kit (QIAGEN). YFP coding sequence in a lentiviral vector (Addgene, 61425) was cloned into the packaging cassette of dCas9-VP64 (Addgene, 61422) and MS2-P65-HSF1 (Addgene, 61423) were subcloned into pAd5 vector (Invitrogen) for virus packaging. Viruses were amplified, desalted, and titered as described above. Three adenoviruses (1:1:1) were delivered into humanized mice intravenously at a total of 5 × 10⁸ PFU/mouse. After 7 days, tissue samples were harvested for further analysis after mice were under food withdrawal for around 5 hours (from ~9:00 am to 2:00 pm).
the column from Simple ChIP Enzymatic Chromatin IP kit (Cell Signaling Technology). The primers used were the same as used for mouse Ptbp1 ChIP analysis (Supplemental Figure 5).

RIP analysis
To prepare liver tissue lysates, frozen liver tissues were homogenized using a Dounce homogenizer with 15-20 strokes in RIP buffer (150 mM NaCl, 20 mM Tris pH 7.4, 1 mM EDTA, 0.5% Triton X-100 with Protease/Phosphatase Inhibitor Cocktail and RNase-Out). For each RIP, 5 μg rabbit IgG or Ptbp1 antibody (32-4800, Thermo Fisher Scientific) were first incubated with 30 μL washed Dynabeads Protein G in 300 μL RIP buffer supplemented with 0.2 mg/mL BSA, 0.2 mg/mL heparin and 0.2mg/mL EcoRI (RNA) for 1 hour. Then, the antibody-coupled beads were added to 5 mg of liver tissue lysates diluted in 500 μL RIP buffer and incubated for 3 hours at 4°C with gentle rotation. Beads were washed briefly 5 times with RIP buffer. At the final wash, one-fifth of beads were used for protein analysis and the rest of the beads were resuspended in 1 mL of TRizol for RNA extraction. Coprecipitated RNAs were isolated and analyzed by RT-PCR.

ChIP analysis
ChIP assays of frozen liver tissues of humanized mice were performed using the SimpleChIP Enzymatic Chromatin IP kit (Cell Signaling Technology) according to the manufacturer’s protocol.

IP was performed using RNA Poly II ChIP-validated antibody (MilliporeSigma, 17-620), Ptbp1 antibody (Thermo Fisher Scientific, 32-4800), or with rabbit IgG as a negative control. For RNA poly II ChIP, the DNA in each ChIP were determined by qPCR analysis using primers amplifying the genomic sequences covering the transcriptional start sites of genes. For Ptbp1 ChIP, the DNA in each ChIP were determined by qPCR analysis using primers amplifying the genomic sequences covering the promoters of genes. The primers used are list in Supplemental Figure 5. The relative enrichment was calculated by normalizing the amount of ChIP DNA to input DNA and comparing with the IgG control as fold enrichment.

Luciferase reporter assay
The human HMGCASI promoter was amplified (forward, GTCCATCGGAATTAGTTAGCTGTGC; reverse, CAATCGCGGCCGGTATGAGTGG) and cloned into the pGL3-Basic Vector (Promega). Full-length PTBPI expression vector and control vector were purchased from OriGene (RC201779 and PS100001). The HEK293A cells were maintained in DMEM supplemented with 10% cosmic calf serum. Cells were transfected with pGL3-HMGCASI promoter, the PTBPI, pAd-hLMR1, or control vectors using Lipofectamine 2000 (Invitrogen), and luciferase assays were performed 24 hours later using the Dual-Luciferase Reporter Assay Kit (Promega). Transfection efficiency was measured by normalization to Renilla luciferase activity expressed from a cotransfected pTK-RL vector (Promega).

Immunoblotting
For immunoblotting analyses, the cells and tissues were lysed in 1% SDS lysis buffer containing phosphatase inhibitors (Sigma-Aldrich) and a protease inhibitor cocktail (Roche). The lysate was subjected to SDS-PAGE, transferred to PVDF membranes, and incubated with the primary antibody followed by the fluorescence conjugated secondary antibody (LI-COR). The bound antibody was visualized using a quantitative fluorescence imaging system (LI-COR). The Ptbp1 antibody (32-4800) was from Thermo Fisher Scientific.

Measurement of lipid levels in liver tissues and plasma
The liver and plasma cholesterol levels in regular mice were measured by using a cholesterol assay kit from Abcam (ab65390) and normalized to tissue weights. To measure human LDL-VLDL cholesterol levels in the plasma of humanized mice, human apolipoprotein B–containing lipoproteins in the plasma of humanized mice were immunoprecipitated by using LipoSep IP reagent (Sun Diagnostics, LS-01). The immunoprecipitated pallets were first washed with PBS, and then were resuspended in PBS plus 0.5% NP-40 to release lipids from lipoproteins. After a brief centrifugation, the cholesterol levels in supernatant were measured by using a cholesterol assay kit from Abcam (ab65390) (as in Supplemental Figure 1D), or further normalized to the humanized ratio of each mouse (as in Figure 4D). The humanized ratio was determined by the relative expression levels of human 16S in the real-time PCR analyses using cDNA prepared from the homogeneous powder of each humanized liver tissue.

Statistics
For comparisons between 2 groups, a 2-tailed, unpaired Student’s t test was used in the following figures: Figure 4, A and D, fasting versus fed, refeeding versus fasting; Figure 5, A–D; Figure 6, A and D; Figure 7, A–C; and Supplemental Figure 4, B and C. For multiple comparisons, 1-way ANOVA with post hoc Tukey’s honestly significant difference test was used in Figure 6C. A P value of less than 0.05 was considered significant.

Study approval
All animal experiments were performed in accordance with and with approval from the NHLBI Animal Care and Use Committee or the Animal Care Committee of the CIEA, Kawasaki, Japan. All human-related data sets were downloaded from public domains.

Author contributions
XR, PL, YM, CJ, and HC designed the workflow. XR, PL, YM, and CJ performed most of the experiments with assistance from XR, PL, YM, CJ, and HC designed the workflow. XR, PL, YM, CJ, and HC performed the bioinformatics analysis. XR, PL, YM, CJ, and HC wrote the manuscript. HC conceived and supervised the study. The order of the co–first authors is based on the length of time spent on and the contribution to the project.

Acknowledgments
We thank James Hawkins (NHLBI Animal Program) for assistance with the maintenance of humanized mice; Yan Luo, Poching Liu, and Yuesheng Li (NHLBI DNA Sequencing and Genomics Core) for RNA-Seq analysis; and Yong Chen and Marjan Gucek (NHLBI Proteomics Core) for their help with mass spectrometry analysis. We also thank Roy Parker and Carolyn Decker (University of Colorado Boulder) for access to the DeltaVision Elite microscope, and Theresa Nahreini and Nicole Kethley for the use of the Cell Culture Facility (University of Colorado Boulder). We are grateful to Michael N. Sack
(NHLBI Cardiovascular Branch) for critical reading of the manuscript and helpful comments. The GTex project was supported by the Common Fund of the Office of the Director of the NIH, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. NG was supported by the NIH Medical Research Scholars Program, a public-private partnership supported jointly by the NIH and contributions to the Foundation for the NIH from the Doris Duke Charitable Foundation (DDCF grant 2014194), Genentech, Elsevier, and other private donors. DDCF funds were not used to sponsor any of the research that involved animals. This study was funded by NHLBI Division of Intramural Research funds to HC (1ZIAHL006103, 1ZIAHL006159).

Address correspondence to: Haiming Cao, Cardiovascular Branch, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA. Phone: 301.402.9032; Email: haiming.cao@nih.gov. NG’s present address is: Cleveland Clinic Foundation, Cleveland, Ohio, USA.