Article

Human genetic variation determines 24-hour rhythmic gene expression and disease risk

Ying Chen[1], Panpan Liu[1], Aniko Sabo[2] & Dongyin Guan [1]✉

24-hour biological rhythms are essential to maintain physiological homeostasis. Disruption of these rhythms increases the risks of multiple diseases. Biological rhythms are known to have a genetic basis formed by core clock genes, but how individual genetic variation shapes the oscillating transcriptome and contributes to human chronophysiology and disease risk is largely unknown. Here, we mapped interactions between temporal gene expression and genotype to identify quantitative trait loci (QTLs) contributing to rhythmic gene expression. These newly identified QTLs were termed as rhythmic QTLs (rhyQTLs), which determine previously unappreciated rhythmic genes in human subpopulations with specific genotypes. Functionally, rhyQTLs and their associated rhythmic genes contribute extensively to essential chronophysiological processes, including bile acid and lipid metabolism. The identification of rhyQTLs sheds light on the genetic mechanisms of gene rhythmicity, offers mechanistic insights into variations in human disease risk, and enables precision chronotherapeutic approaches for patients.

Biological rhythms refer to recurring physiological processes with a periodicity of approximately 24 h. These rhythms allow mammals to anticipate daily environmental changes, including light-dark cycles, and are essential to maintain physiological homeostasis[1–3]. The disruption of biological rhythms is increasingly recognized as a risk factor for multiple diseases, including type 2 diabetes, cardiovascular disease, digestive disease, and cancer[4–6]. It is generally accepted that variations in the 24 hrhythms regarding gene expression and physiological processes exist in different individuals[7,8]. Nevertheless, the major questions remain regarding whether and how these variations in biological rhythms contribute to disease risk, as well as what underlying mechanisms are responsible for these variations.
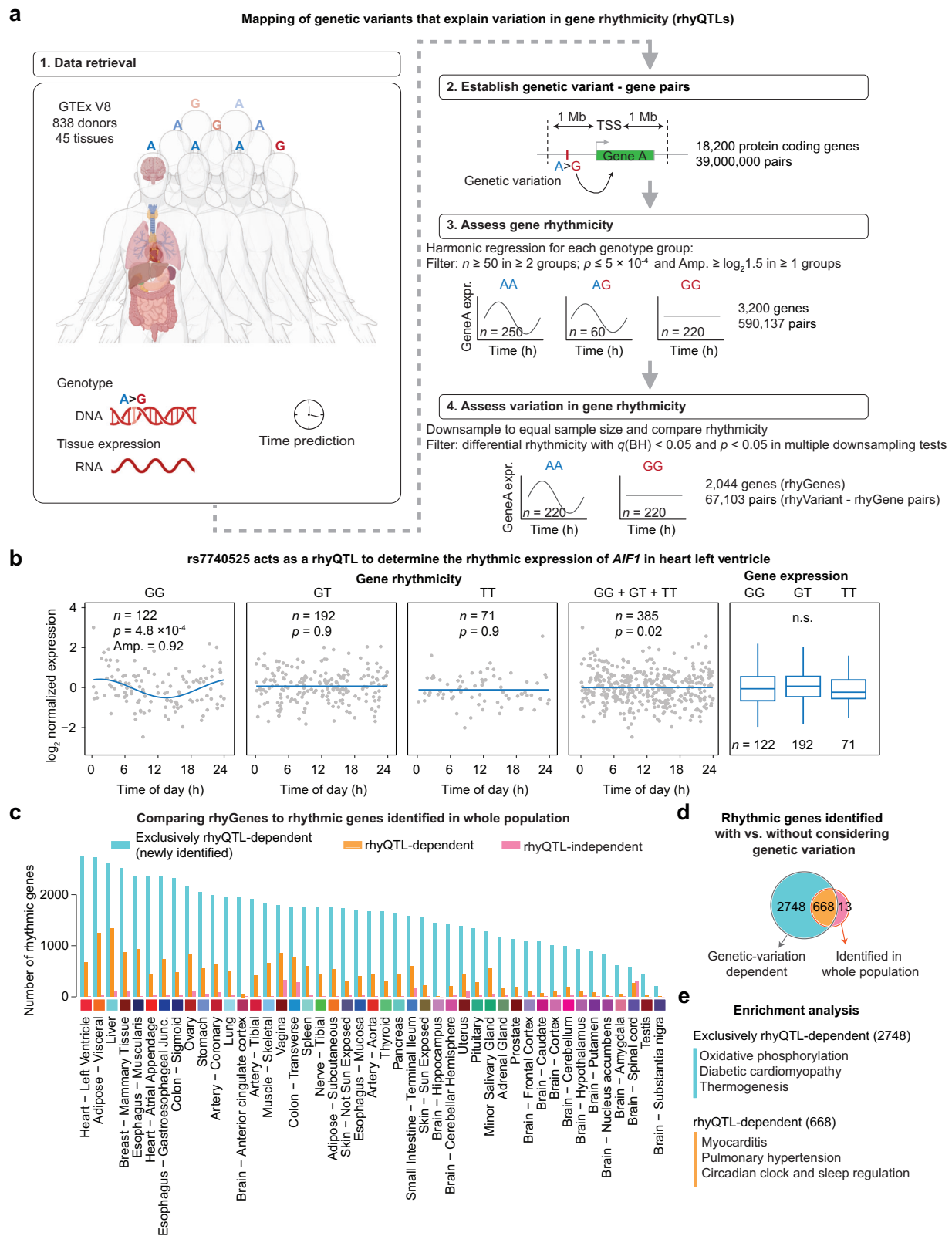
Moreover, our current understanding of mechanisms regulating rhythmic gene expression highlights the regulatory roles of transcription factors (TFs) on *cis*-regulatory elements (CREs), including core clock components such as BMAL1 and REV-ERBs, as well as non-canonical clock TFs[9–11]. Genetic or environmental perturbation of these regulatory TFs in animal models can cause or exacerbate multiple diseases[12,13]. Variants in putative CREs have been linked to variations in human gene regulation, sleep disorders, individual chronotypes, and other complex traits and diseases[14–16]. Although these variants in CREs

may affect core clock gene expression, the specific TFs targeting these CREs and their mechanisms of action remain poorly defined. Moreover, the relationships between genetic variation and human biological rhythms in specific tissues are largely unexplored, as are the mechanisms linking these variations to complex traits and diseases in humans. Since direct genetic manipulation is impractical in humans, we studied natural genetic variation in the Genotype Tissue Expression (GTEx) Project to identify associations with perturbations in gene rhythmicity, which refers to the periodic fluctuation of gene expression with a 24-hour cycle, and to further determine the relationship between gene rhythmicity and human phenotype in various tissues.

Results

We first established a median of 39 million genetic variant-gene pairs across 45 tissues from 838 individuals using GTEx data (Supplementary Data 1)[17] and then assessed the association between genetic variants and rhythmic expression of their nearby genes (Fig. 1a). Using harmonic regression to evaluate the gene rhythmicity in a subpopulation with specific genotype for each genetic variant-gene pair, we identified a median of 3200 genes across 45 tissues that are rhythmically expressed in at least one genotype subpopulation. A

[1]Division of Diabetes, Endocrinology, and Metabolism, Department of Medicine, Baylor College of Medicine, Houston, TX, USA. [2]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. ✉e-mail: dongyin.guan@bcm.edu

a Mapping of genetic variants that explain variation in gene rhythmicity (rhyQTLs)

b rs7740525 acts as a rhyQTL to determine the rhythmic expression of *AIF1* in heart left ventricle

c Comparing rhyGenes to rhythmic genes identified in whole population

d Rhythmic genes identified with vs. without considering genetic variation

e Enrichment analysis

median of 2044, accounting for 63.8% of total rhythmic genes across tissues, exhibits differential rhythmic expression among genotypes (Fig. 1a and Supplementary Fig. 1a–c). For example, the expression of allograft inflammatory factor 1 (*AIF1*) in heart tissue, which correlates with the development of cardiac allograft[18,19], is only rhythmic in the subpopulation with GG genotype at single nucleotide polymorphism (SNP) rs7740525 (Fig. 1b). Note that the overall expression levels of

AIF1 have no difference across these three genotypes. This result indicates the relationship between SNP rs7740525 and *AIF1* rhythmicity cannot be explained by expression quantitative trait loci (eQTLs), which are associated with variations in gene expression abundance across different genotypes. Here, we refer to the genetic loci that are associated with variations in 24-hour rhythmic gene expression as rhythmic quantitative trait loci (rhyQTLs) and the genes associated

Fig. 1 | Genome-wide mapping of rhyQTLs across human tissues. a Study design for rhyQTL mapping and rhyGene identification. The number of genes and gene-genetic variation pairs displayed in the plot represent the median values across 45 solid tissues used in this study. In step 3, the parameters of *p* value and amplitude (log₂ fold change of peak-to-trough) are assessed by harmonic regression fit. In step 4, variation in gene rhythmicity is assessed by harmonic ANOVA (HANOVA). Adjusted *p*-values were further computed with the Benjamini-Hochberg (BH) correction. Created in BioRender. Guan, D. (2025) https://BioRender.com/nlykndj. **b** SNP rs7740525 as a representative rhyQTL determines the rhythmic expression of *AIF1* in the subpopulation with GG genotype but not in other genotypes. The parameters of *p*-value and amplitude by harmonic regression fit are shown. The expression levels of *AIF1* among the three genotypes showed no statistical significance. Box plots represent gene expression levels across three genotypes. The

median (center line), 25th and 75th percentiles (box), and data within 1.5 × inter-quartile range (whiskers) are shown; minimum and maximum values beyond this range (outliers) are not shown. Gene expression differences across genotypes are assessed using linear regression. Sample size (*n*) is labeled in the figure. **c** Overlap between rhythmic genes identified with and without considering rhyQTLs across 45 tissues. The rhythmic genes identified in the whole population only include those located on autosomes, as genes on sex chromosomes were not considered in the identification of rhyGene and rhyQTL. **d** Venn diagram for rhyQTL-dependent rhythmic genes and rhythmic genes identified from the whole population without considering genetic variation in the left ventricle heart. **e** Pathway enrichment analysis on exclusively rhyQTL-dependent and rhyQTL-dependent genes in the left ventricle heart.

with at least one rhyQTL as rhyGenes. In line with previous studies, we identified rhyQTLs that are associated with chronotype[15,20] (Supplementary Fig. 2a). For example, rs10788872 and rs2055975 are linked to individuals who prefer morning activities, such as going to bed and waking earlier[15]. These SNPs are associated with the rhythmic expression of the Circadian Associated Repressor of Transcription (*CIART*) in the brain tissues (Supplementary Fig. 2b, c), which regulate circadian rhythms by modulating the activity of key circadian clock components[20]. This association suggests that changes in the rhythmic expression of *CIART* may reflect the genetic basis of chronotype preferences. We also detected rhyQTLs exhibiting sex-specificity, with 11% and 21% specifically mediating rhythmic gene expression in males and females, respectively, in adipose visceral tissue (Supplementary Fig. 3).

Since the rhythmic expression pattern in a subpopulation with a specific genotype could be masked by the non-rhythmic expression of the same gene in subpopulations with other genotypes, the rhythmic expression of a single gene may only be observed in a subpopulation with a specific genotype but not in the whole population. Indeed, our rhyQTL analysis uncovered a median of 4 times as many rhythmic genes across 45 tissues than previously recognized in whole human populations without consideration of genetic variation[21] (Fig. 1c). For example, in heart tissue, 2748 rhythmic genes are newly identified, displaying rhythmic expression exclusively in specific genotypes. Even in 681 genes whose rhythmicity is detected from whole populations, 98% of them are affected by rhyQTLs, and only 2% of rhythmic genes from the whole population are independent of *cis*-regulation by nearby genetic variants (Fig. 1d). To distinguish the functional importance between genetic variation-dependent and independent rhythmic genes, we performed pathway enrichment analysis. Diabetic cardiomyopathy and oxidative phosphorylation pathways are enriched in rhythmic genes that are specifically associated with rhyQTLs (Fig. 1e). Thirteen genetic variation-independent rhythmic genes include those with functions fundamental to the heart. For example, no rhyQTL was identified for Myosin Light Chain Kinase 3 (*MYLK3*), a kinase exclusively expressed in cardiomyocytes and critical for phosphorylating myosin light chains to ensure proper cardiac muscle contraction[22]. The absence of rhyQTLs for *MYLK3* suggests that its rhythmicity is evolutionarily essential. Genetic variations that disrupt this rhythmicity may have been subject to purifying selection, preserving its stability within the human population. In sum, these data indicate the crucial *cis*-regulatory role of genetic variations in modulating rhythmic gene expression.
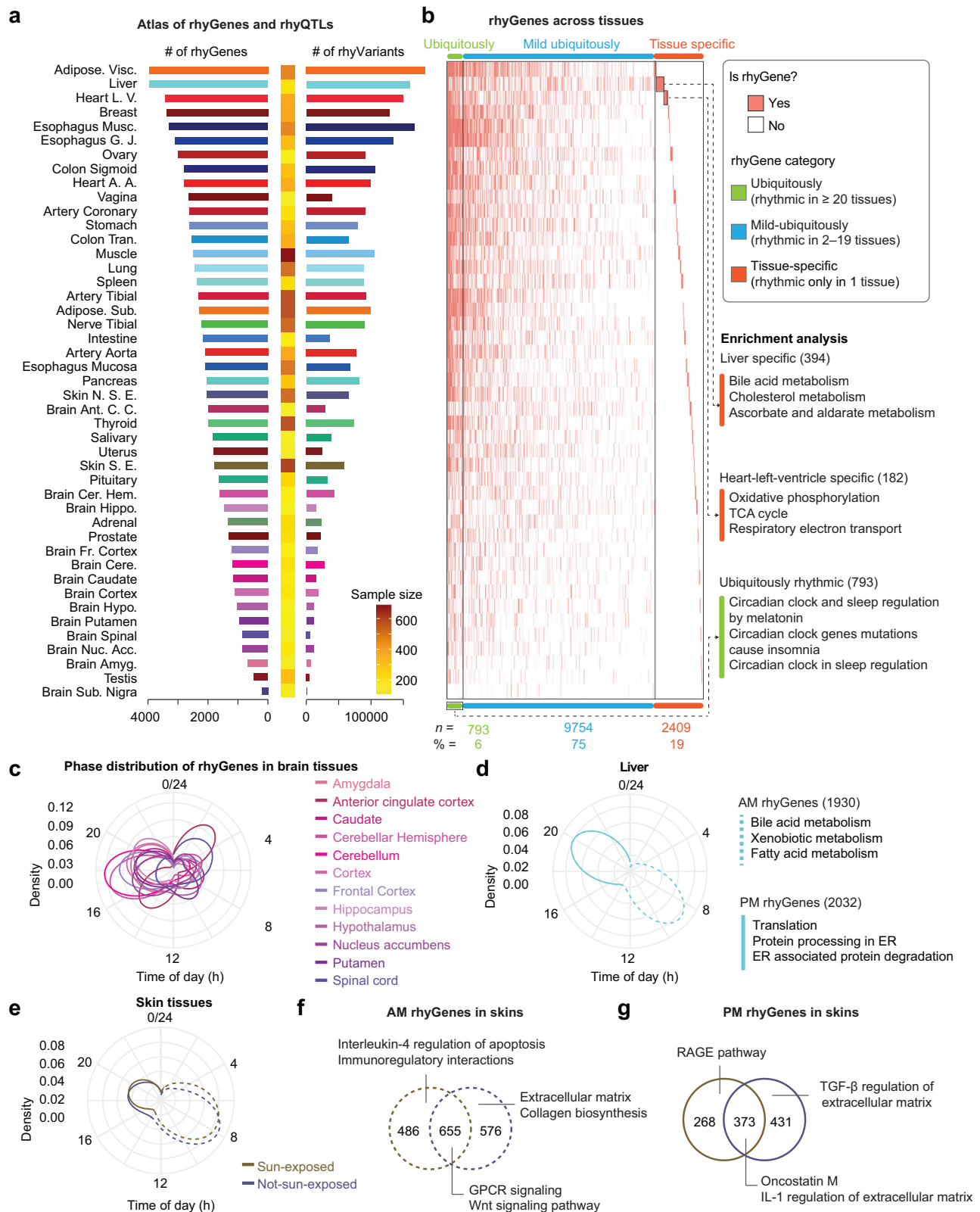
Across different tissues, we observed varying numbers of rhyGenes and their associated variants (rhyVariants), including 3965 rhyGenes and 183,758 rhyVariants in adipose visceral tissue, and 202 rhyGenes along with 653 rhyVariants in substantia nigra (Fig. 2a and Supplementary Data 2). To account for LD and report independent signals, we grouped these rhyVariants into LD blocks, resulting in 6790 independent rhyQTLs in adipose visceral tissue and 210 in substantia nigra (Supplementary Fig. 1c). To evaluate the performance of our rhyQTL detection method with a sample size of 838 individuals, we

conducted two analyses. First, we assessed the power of rhyQTL discovery by randomly selecting subsets of samples of various sizes (Supplementary Fig. 4 and detailed information in "Methods"). Second, we validated the rhyQTLs identified in the GTEx dataset using diurnal RNA-seq data (Supplementary Fig. 5 and detailed information in "Methods"). The results indicate that the datasets from GTEx are effective in detecting genuine genetic variations associated with rhythmic expression variations in humans.

In addition to the *cis*-regulation, we also explored the *trans*-effects of rhyQTLs on rhythmic gene expression. To do so, we focused on core clock genes since their relationship with target genes has been well established in animal models, including core clock gene transcription feedback loops[23]. We first determined whether the rhythmic expression of core clock genes is associated with genetic variations in various tissues. Indeed, the rhythmic expression of core clock components, including *ARNTL*, *NR1D1* and *NR1D2*, are associated with genetic variations in multiple tissues (Supplementary Fig. 6a). For example, SNP rs11870683, which correlates to bipolar disorder in genome-wide association studies (GWAS)[24], contributes to the rhythmic expression of *NR1D1* in the subpopulations with genotype TT (Supplementary Fig. 6b). Next, we evaluated the effects of *NR1D1*-associated rhyQTLs on the rhythmic expression of distal genes located either on the same chromosome or on different chromosomes (Supplementary Fig. 6c). We identified 339 distal genes, including *ARNTL*, *CRY2*, *PER1*, and *PER3*, whose rhythmic expression are tightly correlated with the rhyQTLs of *NR1D1* (Supplementary Fig. 6d and Supplementary Data 3). In sum, these results indicate the *trans*-regulatory role of rhyQTL on distal genes. Given that the *trans*-association is indirect and involves complex regulatory networks, we focused on *cis*-effects of rhyQTLs for high-confidence regulation in the following analyses.

To characterize the genes affected by rhyQTLs, we first quantified the number of tissues where the gene is detected as rhyGene and identified 793 rhyGenes that are common in more than 20 tissues. These common rhyGenes are related to general circadian functions, including melatonin, insomnia, and sleep regulation (Fig. 2b). Interestingly, tissue-specific rhyGenes, defined as rhyQTL-related genes expressed rhythmically in a single tissue but not others, were intricately linked to the distinctive functions of the tissue (Fig. 2b). For example, bile acid and cholesterol metabolism pathways are enriched in liver-specific rhyGenes, while oxidative phosphorylation and tri-carboxylic acid (TCA) cycle pathways are enriched in heart-specific rhyGenes. Notably, the tissue specificity of rhyGenes is not due to limited testing of these genetic variant-gene pairs in other tissues but rather reflects the inherent tissue-specific regulation of rhythmicity (Supplementary Fig. 7).

Next, we explored whether the peak expression of the identified rhyGenes is enriched at specific times of the day using phase analysis. We found that the phase of the rhyGenes is enriched around 5 AM or 5 PM in brain tissues (Fig. 2c and Supplementary Data 4). Coinciding with the concept that brain clocks act as a pacemaker to synchronize biological rhythms in peripheral tissues[25], peripheral tissues show a 2–6 h

a Atlas of rhyGenes and rhyQTLs

b rhyGenes across tissues

c Phase distribution of rhyGenes in brain tissues

d Liver

e Skin tissues

f AM rhyGenes in skins

g PM rhyGenes in skins

delay in their phases (Fig. 2d, e and Supplementary Fig. 8). To determine which physiological processes are regulated by these rhyGenes, we performed pathway enrichment analysis and found that xenobiotic, bile acid, and fatty acid metabolism are enriched in hepatic rhyGenes with peak expression in the morning, while protein translation and protein degradation are enriched in hepatic rhyGenes with peak expression in the afternoon (Fig. 2d). Intriguingly, the phases in

humans regarding metabolism and stress pathways exhibit a 12 h shift compared to these in nocturnal mouse models[26]. These morning and evening waves of rhyGenes are not directly regulated by sunlight exposure because the rhyGenes in sun-exposed and not-sun-exposed skin tissues show similar phase distribution (Fig. 2e). Although the expression waves are similar, the rhyGenes are different in that cell adhesion and metabolism pathways are enriched in not-sun-exposed

skin, while inflammation and immuno-response pathways are enriched in sun-exposed skin (Fig. 2f, g). Together, these results indicate that rhyQTLs are associated with a wide range of rhythmic gene expression in various tissues to regulate time-dependent physiological processes.

To further dissect the potential distinct or combinational effects on gene rhythmicity and abundance contributed by genetic variation, we calculated the proportion of rhyQTL variants (without LD clumping) that are also eQTL variants and found that only 3–37% of rhyQTLs in various tissues are also eQTLs (Fig. 3a). To perform a rigorous comparison with the same model between eQTLs and rhyQTLs, we compared the overall gene expression levels between the same two genotype groups used for rhyQTL identification in each tissue. Our analysis revealed that only 3–12% of the rhyQTLs were also associated with gene expression variations across different tissues (Supplementary Fig. 9). This observation aligns with previous findings demonstrating the decoupling of transcriptomic changes in circadian rhythmicity from mean 24 h gene expression levels[27]. These results indicate rhyQTLs are largely different from eQTLs and majorly regulate gene rhythmicity but not overall expression level. To compare the molecular mechanisms of identified rhyQTLs relative to eQTLs, we performed enrichment analyses using the annotation genomic functional elements defined by VEP[28]. We found that although both eQTLs and rhyQTLs are enriched in various gene regulatory regions (odds ratio > 1), rhyQTLs are more enriched in enhancer regions compared to eQTLs (Fig. 3b and Supplementary Data 5). This is consistent with the findings that clock regulators prefer binding to enhancer regions in mouse models[26,29], suggesting rhyQTLs contribute to variations in enhancer activities for regulating target rhyGenes.

Emerging studies indicate that QTLs convert genetic information to its function by alternating TF DNA-binding consensus sequence and affecting TF chromatin binding affinity[16,30]. To identify which TFs may be regulated by this mechanism, we scanned putative TFs that bind to rhyQTLs by cross-referencing rhyQTLs with the known human TF motifs. We found that binding motifs of core clock genes, such as BMAL1, CLOCK, and NR1D1 were highly enriched in rhyQTLs from brain tissue, aligning with their established roles in rhythmic gene expression. Beyond the core clock components, noncanonical regulators such as Heat Shock Factor 1 (HSF1), which have been reported to exhibit rhythmic DNA binding to heat-shock response genes[31], also showed motif enrichment in these loci (Fig. 3c and Supplementary Data 6). Additionally, the binding motifs of TFs that are functional in brain tissue, such as ELK1, which relates to nervous differentiation[32,33], was found to be enriched in rhyQTLs identified in brain frontal cortex (Fig. 3c). Notably, the enrichment of binding motifs for core clock genes and HSF1 was observed across most tissues (Fig. 3d and Supplementary Data 6), indicating the interactions between genetic variations and clock regulators in regulating the variation of biological rhythms.

Genetic variants, especially those in CREs, have been linked to human complex traits and diseases through GWAS[34–36]. To decipher whether rhyQTLs can explain the variations in human traits and diseases, we retrieved around 500,000 published SNP-trait/disease associations from GWAS Catalog[37] and assessed the enrichment of rhyQTLs within these GWAS-tag SNPs across different tissues (Supplementary Fig. 10a). We found that rhyQTLs contribute to multiple

traits/diseases and show more statistical significance in GWAS than control SNPs (Supplementary Fig. 10b, c and Supplementary Data 7). Interestingly, hepatic rhyQTLs exhibit the highest enrichment among GWAS-tagged SNPs (Fig. 4a). To further dissect the specific traits/diseases that were contributed by hepatic rhyQTLs, we classified the traits/diseases from GWAS Catalog into specific physiological function or disease-related categories and then calculated the enrichment of these rhyQTLs in each category (Supplementary Fig. 10a and Supplementary Data 8). We observed that hepatic rhyQTLs are enriched across all categories of traits/diseases, with the top enriched traits related to lipid/lipoprotein markers and liver enzyme levels, which directly reflect liver-related functions (Fig. 4b).

Heritability refers to the proportion of variation in a population trait that can be attributed to inherited genetic factors[38]. To further explore whether the heritability of lipid/lipoprotein-related traits is attributed to rhyQTLs, we conducted partitioned heritability for 15 lipid or lipoprotein-related traits using stratified linkage disequilibrium score regression (LDSC)[39–41]. Among these traits, rhyQTLs and eQTLs explained a median of 15.8% and 22.9% SNP heritability, respectively (Fig. 4c). Notably, the proportion of heritability attributed to rhyQTLs in certain trait was comparable to that of eQTLs, as observed for high-density lipoprotein (HDL) cholesterol (Fig. 4c). Thus, these results demonstrate that rhyQTLs extensively contribute to human complex traits and diseases, with the most traits being associated with the liver functions.

To estimate the effect of specific rhyQTLs on specific human phenotypes, we identified rhyQTLs that share the same variants with significant GWAS signals in 15 lipid/lipoprotein-related traits using COLOC PP4 statistic[42]. We identified 66 rhyGene-trait associations involving 29 rhyGenes whose rhyQTLs uniquely colocalized with GWAS signals for these traits, whereas their corresponding eQTLs did not exhibit such colocalization (Supplementary Data 9). For example, Leucine-rich repeat-containing G protein-coupled receptor 4 (*LGR4*) has been reported to link the peripheral circadian clock with lipid metabolism in the liver[43,44]. In our analysis, rhyQTLs of *LGR4* colocalized with significant GWAS loci on chromosome 11 associated with the linoleic acid to total fatty acids ratio. In contrast, no strong association was observed between eQTLs of *LGR4* and these GWAS loci (Fig. 4d). Similarly, colocalization patterns were identified between rhyQTLs regulating the rhythmic expression of Euchromatic Histone Lysine Methyltransferase 2 (*EHMT2*) and Selenoprotein N (*SELENON*) and significant GWAS loci associated with LDL cholesterol and total cholesterol levels. However, no such colocalization patterns were observed for their corresponding eQTLs (Fig. 4e, f). *EHMT2* has been reported to suppress cholesterol biosynthesis by maintaining repressive H3K9 methylation marks at the promoter of *SREBF2*, a key regulator of cholesterol biosynthesis[45]. *SELENON*, one of the selenoproteins, may play a role in lipid metabolism, as abnormal levels of selenoproteins have been associated with dysregulated lipid profiles[46]. Thus, rhyQTLs provide insights into GWAS signals that remain unexplained by eQTLs, highlighting their distinct regulatory contributions to human traits.

Recent studies indicated that disruption of 24-hour rhythms can exacerbate metabolic disorders in obese animal models[10,47]. To determine whether our identified rhyQTLs could cause or even magnify the
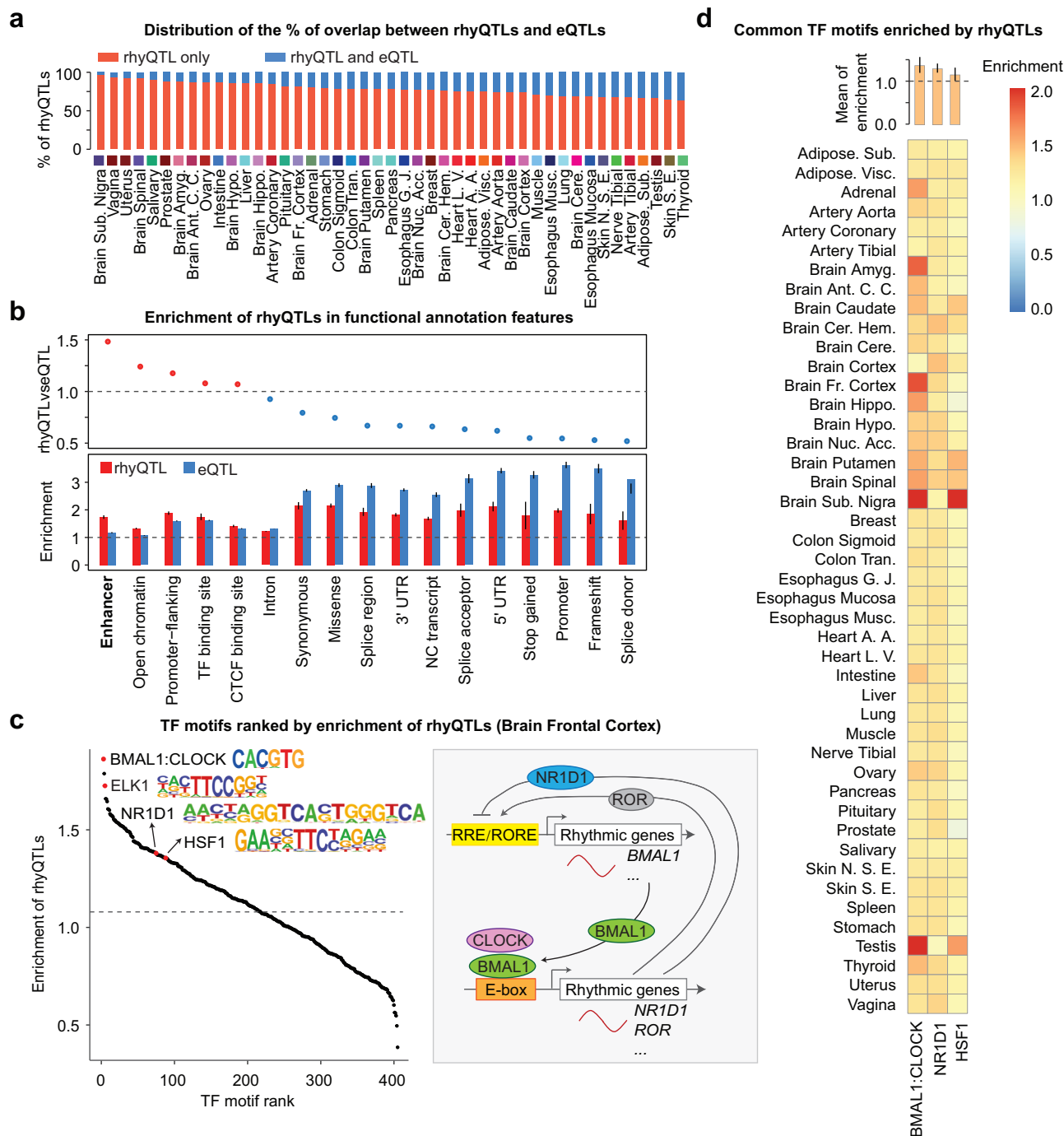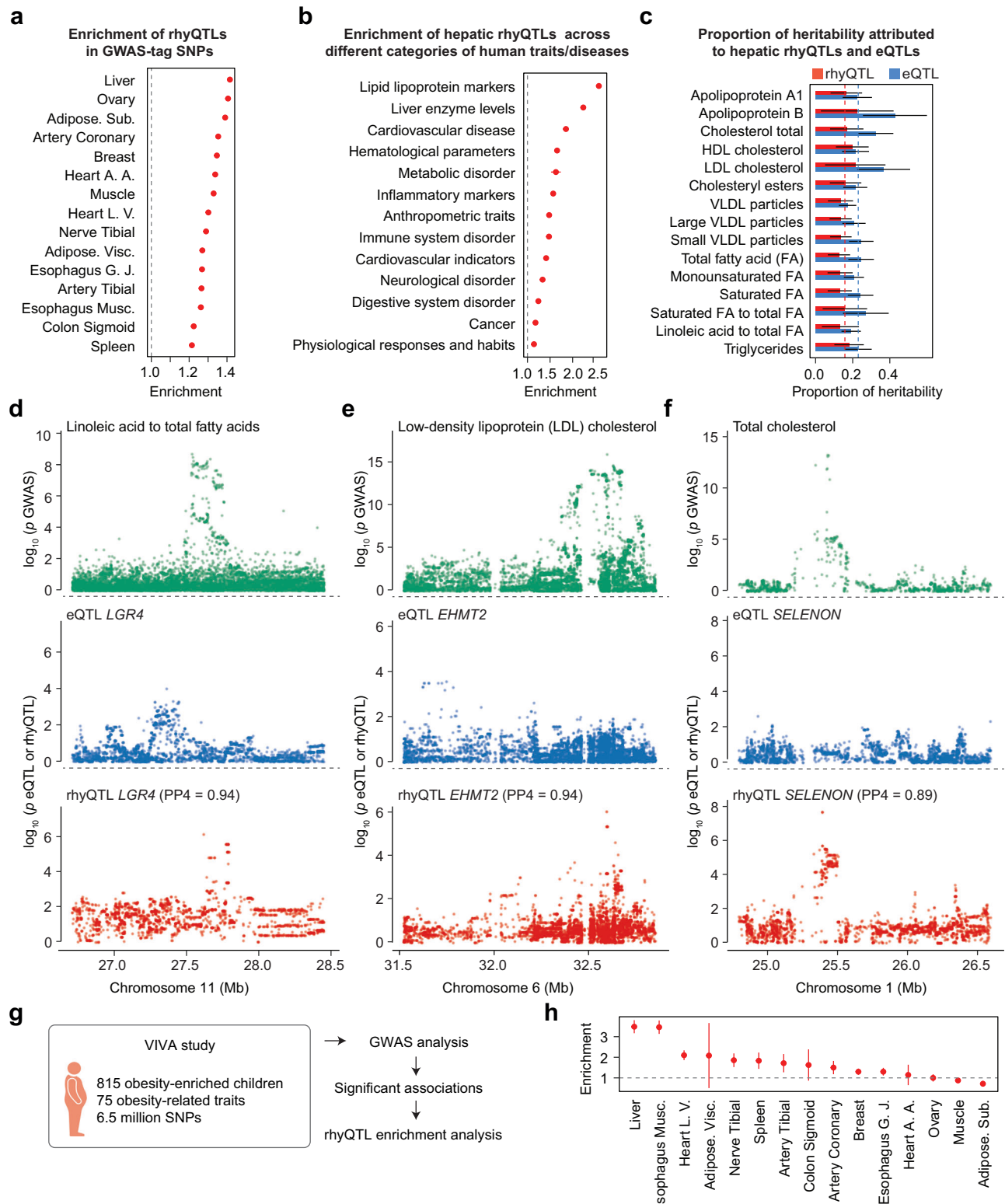
Fig. 3 | rhyQTLs represent a novel type of molecular QTL. a The overlap of genomic loci between rhyQTLs and eQTLs. **b** Enrichment of QTLs in genomic annotation categories defined by Ensembl VEP. Each bar represents the median of enrichment in each category across 45 tissues, with error bars representing the 95% confidence intervals (CIs) across tissues. The enrichment is calculated as an odds ratio score using the numbers of observed QTLs and expected QTLs located in each annotation category compared to those in baseline regions. Baseline regions are unions of all functional and putative functional regions within the human genome. **c** TF motif enrichment in genomic loci harboring rhyQTLs from brain frontal cortex tissue. The enrichment of TF was calculated using 2 × 2 contingency tables. *p*-value was calculated using two-sided Fisher's exact test. The schematic on the right illustrates the circadian clock, a cell-autonomous molecular mechanism driven by transcriptional–translational feedback loops. The clock is driven by the CLOCK:BMAL1 heterodimer, a transcriptional activator that binds to E-box sequences to regulate the rhythmic expression of numerous genes, such as *NR1D1* (also known as *REV-ERBα*) and *ROR* by binding to E-box sequences. In turn, NR1D1 and ROR control the transcription of *BMAL1* and other rhythmic genes through RRE/RORE elements, with ROR acting as an activator and NR1D1 as a repressor. These interconnected cellular circadian oscillators drive the rhythmic expression of circadian output genes. In the scatter plot, each dot represents a TF motif, with core clock components and the TF ELK1, which is functional in the brain highlighted in red. Heat shock factor, known for its rhythmic binding activity, is also highlighted. **d** The heatmap of enrichment of the most enriched motifs across 45 tissues. The bars on the top indicate the mean enrichment across tissues, with error bars representing the standard deviation across different tissues.

a Enrichment of rhyQTLs in GWAS-tag SNPs

b Enrichment of hepatic rhyQTLs across different categories of human traits/diseases

c Proportion of heritability attributed to hepatic rhyQTLs and eQTLs

d Linoleic acid to total fatty acids

e Low-density lipoprotein (LDL) cholesterol

f Total cholesterol

g VIVA study
815 obesity-enriched children
75 obesity-related traits
6.5 million SNPs
→ GWAS analysis
Significant associations
rhyQTL enrichment analysis

h Enrichment

variation in metabolic phenotypes among obese individuals, and to further evaluate their functional relevance in obesity-prone populations, we performed GWAS analysis on 75 obesity-related traits in a cohort containing 815 children enriched for obesity (Fig. 4g and Supplementary Data 10). Our analysis identified 186 SNP-trait associations, including a GWAS signal on chromosome 11 for TNFα levels that was uniquely detected in this specific population but absent in the general population (Supplementary Data 11). SNPs contributing to this signal were identified as rhyQTLs in multiple tissues, particularly in the liver,

where they were associated with the rhythmic expression of *IGF2*, which regulates TNFα signaling and mediates inflammatory responses[48,49]. Moreover, the enrichment analysis revealed that rhyQTLs in multiple metabolic tissues, including liver, colon, and adipose tissues, contribute to the above obesity-related traits (Fig. 4h). These results highlight the potential significance of rhythmic genetic regulation in modulating phenotypic variation relevant to metabolic dysfunctions observed in obesity-prone populations. Together with prior observation linking circadian disruption to exacerbated

Fig. 4 | rhyQTLs contribute to human traits/diseases. a Enrichment of rhyQTLs in GWAS-tagged SNPs. Enrichment was calculated as the ratio of observed SNPs that were both hepatic rhyQTLs and GWAS-tagged SNPs to expected SNPs. The expected SNPs were randomly selected 30 times, with each selection matched in number and minor allele frequency (MAF) to the observed hepatic rhyQTLs. The median ratio across 30 iterations was used to quantify enrichment. The figure displays the top 15 tissues with the highest number of rhyQTLs. **b** Enrichment of hepatic rhyQTLs in the GWAS SNPs associated with different categories of human diseases/traits. **c** Partitioned heritability plot for the proportion of phenotypic variance that can be explained for 15 traits by rhyQTLs and eQTLs. The proportion of heritability is quantified as the ratio of heritability attributed to rhyQTLs or eQTLs to the overall SNP-based heritability using stratified linkage disequilibrium score regression (LDSC). The red and blue dashed lines represent the median values of rhyQTL and eQTL across traits, respectively. Error bars indicate the 95% CIs of the

estimates shown in (**a**–**c**). **d**–**f** Representative examples demonstrate that traits are associated with rhyQTLs rather than eQTLs. The top track shows the -$\log_{10}(p)$ values of SNPs from the GWAS of linoleic acid to total fatty acids (**d**), LDL cholesterol (**e**), and total cholesterol (**f**) levels. The middle and bottom tracks display the -$\log_{10}(p)$ values in the GWAS of eQTL SNPs (middle track) and rhyQTL SNPs (bottom track) within the region, respectively. Colocalization between rhyQTLs (or eQTLs) and GWAS loci is assessed using *coloc*. Colocalizations with a posterior probability for the shared variants (PP4) > 0.70 are considered significant. **g** The schematic illustrates the GWAS analysis conducted in the VIVA study. Created in BioRender. Guan, D. (2025) https://BioRender.com/r4h4w79. **h** Enrichment of rhyQTLs among significant GWAS SNPs identified in an obesity-enriched cohort from VIVA study. The top 15 tissues with the highest number of rhyQTLs are shown. Error bars indicate the 95% CIs of the estimates.

metabolic disorders in obese animal models, our findings provide human population-level insights into how rhythmic genetic variations may shape individual susceptibility to obesity-related phenotypes.

Discussion

By analyzing genome-wide genotypes and nearby gene rhythmicity, we mapped a new type of QTL, termed rhyQTL, that determines the variations in rhythmic gene expression in various human tissues. Besides *cis*-effects, we also detected the *trans*-effects of rhyQTLs by using core clock gene *NR1D1*-associated rhyQTLs and indicated the *trans*-regulatory role of rhyQTLs on distal gene rhythmicity (Supplementary Fig. 6). rhyQTLs provide novel insights into the variation in chronophysiology among individuals, such as bile acid and cholesterol metabolism. The genomic loci of rhyQTLs are largely different from eQTLs in terms of genomic coordinates and regulatory mechanisms. Compared with eQTL, rhyQTLs are more enriched in enhancer regions, which fits the general concept that the gene rhythmic expression is interdependent on environmental cues that mediate enhancer activity and intrinsic clock machine[9,26,29]. Regarding the functional importance of rhyQTLs and variations in gene rhythmicity, we have observed that rhyQTLs (independent of eQTLs) extensively contribute to variations in human disease risk, such as metabolic disorders and cardiovascular diseases. Taken together, rhyQTLs-mediated variations in rhythmic gene expression provide a different angle to explain GWAS signals, additional considerations on genetic variation for understanding chronophysiology in humans, and rationales for optimizing chronotherapy based on patients' genetic backgrounds.

Given the potential for sex-specific regulatory mechanisms in gene expression, we also explored the sex-specificity of rhyQTLs. Our analysis focused on adipose visceral tissue, which had a sufficient sample size for meaningful analysis. 11% of rhyQTLs in this tissue were specific to males, while 21% were specific to females (Supplementary Fig. 3). These results indicate a sex difference regarding the impact of rhyQTLs on rhythmic gene expression. Expanding this analysis to multiple tissues with statistically sufficient sample sizes will offer a comprehensive understanding of sex-specific rhyQTL landscapes in the future. Studies will also be needed to dissect the molecular mechanisms underlying these sex-dependent effects.

Despite the valuable insights our study provides, there are some limitations to consider. Unlike traditional QTL analyses, such as eQTL studies that provide a single beta value to represent the linear effect of an SNP, rhyQTL analyses lack effect size estimates (e.g., beta values or Z-scores). To evaluate rhythmicity differences, methods like HANOVA assess multiple dimensions (e.g., amplitude and phase), offering a more comprehensive perspective on how genetic variation modulates rhythmic expression patterns, which cannot be simplified to a single linear effect. As a result, the contribution of SNPs to rhythmicity is inherently multidimensional. It cannot currently be represented by an integrated effect size, such as beta, which does not apply to rhyQTLs. As a result, fine-mapping tools like SuSiE[50,51], which are effective for

separating association signals and estimating causal probabilities for SNPs, are not directly applicable to rhyQTLs. These tools rely on effect size estimates, which are unavailable in rhyQTL analyses and limit their utility for identifying distinct causal SNPs. To address LD and report independent signals, we partitioned genetic variations into independent LD blocks, defining rhyQTLs as genomic loci containing at least one rhyVariants. However, future advancements in statistical methods to estimate standard errors for each rhythmic parameter will be essential for refining effect size calculations and improving the identification of causal variants.

Methods

Data retrieval and data preprocessing for rhyQTL mapping (Step 1 in Fig. 1a)

The genotype data of 838 individuals in GTEx project were downloaded from dbGaP with accession number phs000424.v8.p2. To identify rhyQTLs, we used PLINK[52], a tool set for whole-genome association studies, to filter genetic variants with standard quality control criteria[53]. We retained genetic variants with a minor allele frequency (MAF) ≥ 0.01, Hardy-Weinberg Equilibrium (HWE) $p \geq 10^{-6}$, and located on autosomes. To gain high-confident rhythmic gene expression with sufficient samples, tissues with more than 100 samples were retained, as indicated in Supplementary Data 1. To account for batch effects and population structure in rhyQTL mapping across tissues, we estimated tissue-specific covariates for gene expression levels. Following the instructions of the GTEx Consortium, we included a minimal set of covariates for QTL analysis, including five genotype principal components (PCs), sequencing platform, library construction protocol, and donor sex[17]. Additionally, we incorporated ischemic time, age, and type of death, which were used for rhythmic gene identification across the entire population without considering genotype effects[21]. These covariates were regressed using a linear regression model to minimize confounding effects and improve the accuracy of the analysis. The time in this study is defined as the donor's internal circadian phase rather than the post-mortem sample collection time. This internal phase integrates the relative circadian phases of tissue samples from the same donor based on the assumption that these phases are conserved across donors and are independent of the exact time of death[17]. Importantly, neither the time of death nor the time of sample cryopreservation alone is sufficient to define circadian rhythmicity, as factors such as donor chronotype, time zone position, and cause of death can obscure the true circadian phase[21,54,55]. The genomic annotation data (gencode.v26.GRCh38.genes.gtf) in which isoforms were collapsed to a single transcript per gene based on the GENCODE 26 annotation were downloaded from GTEx portal (https://gtexportal.org/home/).

Establish genetic variant-gene pairs and assess the rhythmicity for rhyQTL mapping (steps 2 and 3 in Fig. 1a)

To determine the *cis*-regulatory effects of genetic variation on gene rhythmic expression, *cis*-genetic variants within the transcription start

site (TSS) ±1 Mb flanking region of each of the protein-coding genes were used to establish the genetic variant-gene pairs. This 1 Mb distance is commonly employed to identify diverse types of molecular *cis*-QTLs[17,56,57]. We established a median of 39 million genetic variant-gene pairs across 45 tissues with approximately 2.9 thousand *cis*-genetic variants per gene. Moreover, for robust statistical analysis, genetic variant-gene pairs with sample sizes greater than 50 in two or three subpopulations with different genotypes were retained to evaluate the gene rhythmicity in each genotype. Gene rhythmicity was determined using harmonic regression (or cosinor regression), a classic nonlinear regression approach to fit sinusoidal functions to time series data and enable the modeling of periodic fluctuations in gene expression[17,58]. The model is represented by the following equation:

$$E_{xy} = \mu_{xy} + a_{xy}\cos\left(\frac{2\pi t_y}{T}\right) + b_{xy}\sin\left(\frac{2\pi t_y}{T}\right) + \varepsilon \qquad (1)$$

in which E_{xy} is the expression level of gene x in donor y; μ_{xy} represents the mean of the expression level of gene x with genotype y; a_{xy} and b_{xy} are the coefficients of the cosine and sine terms, respectively; t_y is the time point of donor y; T is the period which is 24 h in this study; ε is the random error (noise) of the series about the period component. The amplitude and phase of the rhythmicity were calculated as follows:

$$amp = 2\sqrt{a_{xy}^2 + b_{xy}^2} \qquad (2)$$

$$phase = \frac{T}{2\pi}atan2\left(b_{xy}, a_{xy}\right)modT \qquad (3)$$

Here, *amp* represents the fold change from the expression peak to trough, while *phase* indicates the timing of the peak expression. Rhythmicity was further assessed using a likelihood ratio test to compare the harmonic (periodic) model to a flat model ($a_{xy} = b_{xy} = 0$), which assumes no rhythmic variation in expression levels. The significance of the periodicity was evaluated by calculating the *p*-value using the *lrtest* function in *R*.

For each gene, we conducted multiple tests across all its *cis*-genetic variants. To identify the union set of potential rhyVariants and rhyGenes, we estimated the effective number of independent variants (M_{eff}) associated with each gene by accounting for linkage disequilibrium. Adjusting *p* values based on M_{eff} is an effective method to reduce false positives in genetic association studies[59–61]. The median M_{eff} per gene across all analyzed tissues was 15, and accordingly, we established a significant threshold of 5×10^{-4}, which corresponds to an approximate value of 0.01/15. To evaluate the robustness of this threshold, individuals' genotype data were randomly shuffled, and the rhyQTL mapping analysis was performed on the shuffled dataset. The number of rhyVariants identified at the threshold of 5×10^{-4} was then assessed and revealing that the proportion of rhyVariants detected in the shuffled dataset did not exceed 5% of those identified in the real dataset (Supplementary Fig. 1b), supporting the appropriateness of this threshold. Therefore, if genes met the criteria of *p*-value $\leq 5 \times 10^{-4}$ and a fold change of peak-to-trough ≥ 1.5 in at least one genotype, the gene-genetic pair was retained as a union of possible rhyVariants and rhyGenes.

Assess variations in gene rhythmicity for rhyQTL mapping (step 4 in Fig. 1a)

To map the genetic variants that associate with variations in gene rhythmicity, a Bayesian information criterion (BIC) based model-selection algorithm implemented in the *dryR* package, specifically using the *drylm* function[62]. This function is designed to assess differential rhythmicity in time-series data of normally distributed variables and enables the distinction of the differential rhythmicity of genes in

each genetic variant-gene pair across different genotypes. To mitigate biases associated with sample size, a downsampling approach was conducted to ensure equal sample sizes across compared genotypes. To minimize the loss of sample size during downsampling, the top two genotype groups with the most samples were included to assess the differential gene rhythmicity. Specifically, for each genetic variant-gene pair, the sample size for each of the three genotypes (for example, SNP rs6457301 has three genotypes: TT, TC, and CC) was determined, and the top two genotype groups with the largest sample sizes were selected. The genotype group with the more samples was sub-sampled to match the sample size of the smaller group, ensuring equal sample sizes between the two compared genotypes. Subsequently, the differential rhythmicity was evaluated using *drylm* function from *dryR*[62], which provides a fitted model to categorize the differentiation in rhythmicity between the two compared genotypes. All parameters for 24 h rhythms, including rhythmicity, phase shift, and amplitude change, were considered for evaluating the association between genetic variants and rhythmic gene expression. To mitigate bias related to single-time sampling, we performed twenty times of downsampling for *dryR* analysis and then used *G*-test to determine if the frequency of the fitted model provided by *dryR* across the twenty times significantly deviated from the expected frequency.

We further use harmonic ANOVA (HANOVA) to quantify differential rhythmicity between genotype groups. HANOVA models periodicity by fitting sine and cosine wave components to rhythmic expression data[63]. The core idea is to test whether the rhythmic parameters, such as amplitude and phase, differ significantly between genotype groups. By using ANOVA, HANOVA compares a null model (H_0) where both groups share the same rhythmicity parameters to an alternative model (H_1) where the groups have different rhythmicity parameters. This analysis assigns a *p*-value to assess whether rhythmicity is significantly modulated by genetic variation. Adjusted *p*-values were further computed with the Benjamini-Hochberg (BH) correction. Thus, a genetic variation was defined as a rhyVariant based on the following criteria: (1) its paired gene exhibited a rhythmic expression pattern in at least one genotype subpopulation with a peak-to-trough ratio ≥ 1.5 and *p*-value $\leq 5 \times 10^{-4}$; (2) rhythmic expression patterns of its paired gene vary across genotypes assessed by *dryR* and the observed differential rhythmicity was significantly different from the expectation in *G*-test ($p < 0.05$); (3) the differential rhythmicity met the threshold of q(BH) < 0.05. The rhyVariants were partitioned into LD blocks to quantify the number of rhyQTLs, each defined as a genomic locus containing at least one rhyVariant, based on independent LD blocks in human populations[64].

Evaluate the performance of rhyQTLs mapping strategy using downsampling analysis

To evaluate the performance of our rhyQTL detection method with a sample size of 838 individuals, we compared the power of rhyQTL and rhyGene discovery across various sample sizes by randomly selecting samples in adipose, which contains the highest number of rhyQTLs. Samples were randomly selected from 469 adipose visceral tissue samples to create 8 data subsets. The sample sizes in each subset range from 100 to 450 at intervals of 50. In each data subset, genetic variation-gene pairs are established, and gene rhythmicity is assessed using the same methods as in "Establish genetic variant-gene pairs and assess the rhythmicity for rhyQTL mapping" section. To evaluate and mitigate biases associated with sample size, variations in gene rhythmicity across different genotypes were assessed both with and without employing a downsampling strategy. We found that when the sample size exceeded 250, there were mild changes in the numbers of identified rhyVariants and rhyGenes, which suggests that the current sample size can robustly detect rhyVariants and rhyGenes.

Evaluate rhyQTLs identified in the GTEx dataset using diurnal RNA-seq data

To validate and evaluate our rhyQTL detection method with a sample size of 838 individuals, we compared the rhyQTLs identified from GTEx samples with publicly available diurnal RNA-seq data. This dataset was derived from vastus lateralis muscle biopsies collected from 10 healthy donors every 4 h across a day[65]. The temporal expression levels at pre-mRNA (intronic signal) and mRNA (exonic signal) were quantified, and gene rhythmicity was estimated using harmonic regression across 10 donors. Genes exhibiting rhythmicity at both pre-mRNA and mRNA levels in at least two donors ($p < 0.05$ and a peak-to-trough fold change > 1.5) were included in the set of potential rhythmic genes in this cohort. To evaluate rhyQTLs identified in GTEx dataset, genetic variations that occurred in these 10 donors in the transcription region were called based on the reads in RNA-seq data using the Genome Analysis Toolkit (GATK)[66], following the pipeline designed for germline SNP calling from RNA-seq data (https://github.com/gatk-workflows/gatk4-rnaseq-germline-snps-indels). Due to the small sample size of this cohort, genetic variations with a sample size greater than two in at least two genotype groups were used to intersect with rhyVariants identified from the GTEx muscle samples. As a result, we identified a total of 3414 rhyVariant-rhyGene pairs, comprising 3369 SNPs and 269 associated rhyGenes, for the following comparison between the two cohorts. We observed 1970 rhyVariant-rhyGene pairs, comprising 1958 rhyVariants and 234 associated rhyGenes, exhibit differential rhythms regarding rhythmicity (p-value < 0.05), amplitude (fold change > 1.5) or phase (peak phase > 3 h)[67] between top two genotypes with the most samples. Therefore, 58% of the rhyQTLs and 87% of the rhyGenes could be validated in this 10-individual cohort. Notably, among the 3369 rhyVariants, 33% of them were associated with rhyGenes that did not exhibit rhythmicity among all three genotypes in the validation dataset. The lack of rhythmicity of these genes could be due to the limited number of samples. Considering the nature of the validation dataset with 10 donors, the proportion of validated rhyVariants and rhyGenes could be higher with a larger sample size. This also highlights the power of GTEx data for exploring the interactions between genetic variation and rhythmic gene expression within the currently available datasets.

Estimate *trans*-effect of rhyQTLs on rhythmic gene expression

To explore *trans*-effects of rhyQTLs on rhythmic gene expression, we established the genetic variants and gene pairs between rhyQTLs associated with *NR1D1* rhythmic expression and all protein-coding genes. To exclude potential *cis*-effects of these rhyQTLs, we used only *trans*-genetic locus and gene pairs with a genomic distance greater than 1 Mb between the locus and the gene TSS for downstream analyses. For each pair, assessment of gene rhythmicity and variation in rhythmicity across different genotypes was conducted using methods outlined in the above *cis*-rhyQTL mapping section. In total, we identified 339 *trans*-rhyGenes whose rhythmicity could be mediated by *NR1D1*-associated rhyQTLs in 13 brain tissues.

Pathway enrichment analysis

Pathway enrichment analysis was performed using the *R* package *enrichR*, which interfaces with the Enrichr database (https://maayanlab.cloud/Enrichr/)[68]. For the enrichment analysis of rhyGenes identified in specific tissues, all genes expressed in the corresponding tissue and used for rhyQTL mapping were employed as the background gene list. Expressed genes were defined as those with an average count greater than 10 across all samples. All expressed genes across all tissues were used as the background for the enrichment analysis of ubiquitous rhyGenes, which were identified in more than 20 tissues (Fig. 2b). In the comparison of common AM and PM rhyGenes between sun-exposed and non-sun-exposed skin tissues, the background included all expressed genes from these two tissues. Terms with adjusted p-values smaller

than 0.05 in the databases: BioPlanet_2019, Elsevier_Pathway_collection, MsigDB_hallmark_2020, and Elsevier_Pathway_collection were shown.

Enrichment of rhyQTL and eQTL for genomic regulatory elements

The genomic annotation categories defined by the Variant Effect Predictor (VEP) for each SNP were downloaded from GTEx portal. QTL (rhyQTL and eQTL) enrichment in each annotation category was calculated referring to the methodology outlined by Rozowsky et al. [69]. In brief, an odds ratio score was calculated based on a 2 × 2 contingency table, including the numbers of observed QTLs and expected QTLs located in the element compared to those in the baseline regions.

$$Enrichment = \frac{E_{Obs}/E_{Exp}}{B_{Obs}/B_{Exp}} \qquad (4)$$

In which E_{Obs} and E_{Exp} are the number of observed QTLs and the expected QTLs in the annotation category, respectively; B_{Obs} and B_{Exp} are the number of observed QTLs and the expected QTLs in the baseline region. The baseline regions encompass a comprehensive union of all functional and putative functional regions within the human genome. These functional regions include coding regions, untranslated regions, non-coding RNA genes, open chromatin regions, TF binding sites, as well as active and repressed histone peaks derived from various tissue and cell types, along with evolutionarily conserved regions[39]. To generate the expected set of QTLs, we use number-matched randomly selected SNPs with the same MAF distribution as observed QTLs in each tissue. Random sampling was repeated 30 times, and the median value of the odds ratios from these 30 iterations is considered as the enrichment value for each tissue.

Transcription factor (TF) DNA binding motif scan on genomic loci harboring rhyQTLs

We collected 436 human TF motifs from Homer database[70]. The occurrence of each motif in human genome was scanned using FIMO (p-value < 10^{-4})[71] and then intersected with genetic variation loci identified from GTEx dataset using bedtools[72]. These preprocesses established the link between the co-occurrence of TF binding motif and genetic variants at various genomic coordinates. These genetic variants were further classified by whether they were rhyQTLs and whether they were within the motif to generate a 2 × 2 contingency table for each motif. The odds ratio was used as a measure of rhyQTL enrichment in a motif, and Fisher's exact test was applied for statistical significance ($p < 0.05$).

GWAS enrichment analysis

All SNP-trait/disease associations identified in GWAS were obtained from the GWAS Catalog[73] containing 569,163 associations. To generate a set of high-quality SNPs associated with human traits/diseases, we performed the following quality controls by removing insignificant associations (p-values > 5×10^{-8}), associations obtained from non-European studies, and SNPs in the human leukocyte antigen locus (for hg38: chr6:29,723,339−33,087,199). After filtering, we got a clean list of 242,822 SNP-trait/disease associations in GWAS. To assess the enrichment of rhyQTLs in GWAS-tagged SNPs, we generated an expected set by randomly selecting SNPs matched in both number and minor allele frequency (MAF) distribution to the rhyQTLs in each tissue. Enrichment was calculated as the ratio of the number of SNPs that were both rhyQTLs and GWAS-tagged SNPs to the number of SNPs that were both expected rhyQTLs and GWAS-tagged SNPs. This process was repeated 30 times, and the median ratio across these iterations was used to quantify the enrichment of rhyQTLs in GWAS-tagged SNPs.

To explore which traits or diseases are contributed by hepatic rhyQTLs, we evaluated the enrichment of rhyQTLs among these

GWAS-tag SNPs. Traits/diseases were divided into different categories according to their parent terms as per the annotation in GWAS Catalog (Supplementary Data 8). The set of GWAS-tag SNPs in each category was extended by including the SNPs in high linkage disequilibrium (LD scores = 1) with the tag SNPs to ensure more comprehensive coverage of potential causal variants and increase the statistical power in the enrichment analysis. The enrichment of hepatic rhyQTLs in GWAS-tagged SNPs within each trait or disease category was then calculated by comparing observed hepatic rhyQTLs to randomly selected SNPs matched for number and MAF, using the same method as described above.

Stratified linkage disequilibrium score (LDSC) regression

The stratified LDSC regression was used to quantify the heritability attributable to rhyQTLs and eQTLs for the 15 lipid or lipoprotein-related traits[40,41,74,75]. This approach regresses chi-square statistics from the GWAS summary statistics, which were downloaded from GWAS Catalog, with LD scores to estimate partitioned heritability in a disease-specific manner. To do this, binary annotation for rhyQTLs and eQTLs was created, respectively. This annotation assigns a value of 1 to rhyQTL or eQTL SNPs and a zero value to the remaining SNPs in the baseline regions[39]. The LD scores of the rhyQTL and eQTL were computed using SNP genotype data of the individuals of European ancestry from the 1000 Genomes Project Phase 3 with a window size of 1 centimorgan (cM). The proportion of heritability is quantified as the ratio of heritability attributed to rhyQTLs or eQTLs to the overall SNP-based heritability.

Colocalization of QTL with GWAS signal

We used a Bayesian colocalization approach to identify GWAS signals that might exhibit shared genetic effects between rhyQTLs and eQTLs using the *coloc* v5.2.3 *R* package[51,76]. Full summary statistics for 15 lipid- or lipoprotein-related traits were obtained from the GWAS Catalog. As defined by the coloc method, five posterior probabilities (PPs) were calculated: PP0 represents the null model of no association, PP1 and PP2 represent the probabilities that causal genetic variants are associated with disease signals only or rhyQTLs only, respectively, PP3 represents the probability that the genetic effects of disease signals and rhyQTLs are independent, and PP4 represents the probability of colocalization. The probability of one causal variant associated with both traits (PP4) was used to identify significant (PP4 > 0.70) colocalizations.

GWAS on Viva La Familia study (VIVA) dataset

The VIVA was designed to investigate genetic and environmental factors affecting obesity and its comorbidities in Hispanic children. Each family involved in the VIVA cohort was ascertained on a proband with obesity between the ages 4–19 years. The VIVA cohort was highly enriched for obesity, with a high prevalence of elevated body mass index (BMI): most of the parents were either classified as overweight (34%) or obese (57%), and 52% of the enrolled children were classified as obese (above the 95th BMI percentile). Among the obese children, 62% were above the 99th BMI percentile, indicating severe obesity. All participants were genotyped using marker assays included on the Illumina HumanOmni1-Quad v1.0 BeadChips. Subjects and study procedures are described in the VIVA study[77-79]. Briefly, 75 obesity-related phenotypes were measured, including birth weights retrieved from Texas birth records, anthropometric and body-composition traits via dual-energy x-ray absorptiometry, dietary assessments through 24 h recalls, total energy expenditure and substrate utilization monitored using 24 h room calorimetry, physical activity tracked via accelerometry, and fasting biochemistries analyzed using standard techniques (listed in Supplementary Data 10).

GWAS on these 75 obesity-related phenotypes was conducted following the GWAS standard pipeline[80,81]. Firstly, data quality control

was performed to remove SNPs and individuals with insufficient genotyping quality using PLINK v1.90b6.16 64-bit[52]. Specifically, the SNPs with call rates < 98% or with MAF < 0.05 and those with genotypes not in accordance with the Hardy–Weinberg equilibrium ($p > 10^{-6}$) were eliminated. Individuals with a call rate of less than 98% were also excluded. Then, principal component analysis (PCA) was performed to check for population stratification using PLINK. The first 10 principal components, family structure, age, and sex information, were included as covariates in the association study. Here, the primary aim of our GWAS on these 75 obesity-related traits was to independently identify SNPs associated with each trait, not to identify shared or combined effects across traits. We applied the genome-wide significance threshold adjustment using Bonferroni correction for the number of SNPs tested per trait, and additional correction for the number of traits was not applied to avoid an overly stringent threshold and a loss of power in detecting true associations. A genome-wide significance threshold of p was established at $\leq 7.7 \times 10^{-8}$ based on the number of SNPs (644,251 SNPs) included in the association analysis, and SNPs exceeding this threshold were considered as candidate variants (Supplementary Data 11). The enrichment of rhyQTLs in GWAS-tag SNPs identified in populations enriched for obesity was calculated using the same methods as described above in the section of GWAS enrichment analysis.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw genomic sequencing data from the GTEx project V8 are available in the database of dbGaP with accession number phs000424.v8.p2 [https://www.ncbi.nlm.nih.gov/gap/]. The gene expression data are available for download from the GTEx portal: https://www.gtexportal.org/home/downloads/adult-gtex/bulk_tissue_expression. The time information of 838 individuals in the GTEx cohort is available from Zenodo at https://doi.org/10.5281/zenodo.7215362. GWAS summary statistics used in this study were obtained from GWAS Catalog [https://www.ebi.ac.uk/gwas/]. The summary statistics of rhyQTLs generated in this study are provided in Supplementary Data 2. Source Data is provided in this paper. Source data are provided with this paper.

Code availability

The custom source codes to perform the data analysis relevant to this study are available at https://github.com/YingChen10/rhyQTL[82].

References

1. Guan, D. & Lazar, M. A. Interconnections between circadian clocks and metabolism. *J. Clin. Investig.* **131**, e148278 (2021).
2. Liu, Y. & Dekker, J. CTCF-CTCF loops and intra-TAD interactions show differential dependence on cohesin ring integrity. *Nat. Cell Biol.* **24**, 1516–1527 (2022).
3. Allada, R. & Bass, J. Circadian mechanisms in medicine. *N. Engl. J. Med.* **384**, 550–561 (2021).
4. Swanton, C. et al. Embracing cancer complexity: hallmarks of systemic disease. *Cell* **187**, 1589–1616 (2024).
5. Ruan, W., Yuan, X. & Eltzschig, H. K. Circadian rhythm as a therapeutic target. *Nat. Rev. Drug Discov.* **20**, 287–307 (2021).
6. Zielinska-Dabkowska, K. M., Schernhammer, E. S., Hanifin, J. P. & Brainard, G. C. Reducing nighttime light exposure in the urban environment to benefit human health and society. *Science* **380**, 1130–1135 (2023).
7. Lane, J. M. et al. Genetics of circadian rhythms and sleep in human health and disease. *Nat. Rev. Genet.* **24**, 4–20 (2023).

8. Gentry, N. W., Ashbrook, L. H., Fu, Y. H. & Ptacek, L. J. Human circadian variations. *J. Clin. Investig.* **131**, e148282 (2021).

9. Guan, D. et al. Diet-induced circadian enhancer remodeling synchronizes opposing hepatic lipid metabolic processes. *Cell* **174**, 831–842 e812 (2018).

10. Guan, D. et al. The hepatocyte clock and feeding control chronophysiology of multiple liver cell types. *Science* **369**, 1388–1394 (2020).

11. Lundell, L. S. et al. Time-restricted feeding alters lipid and amino acid metabolite rhythmicity without perturbing clock gene expression. *Nat. Commun.* **11**, 4643 (2020).

12. Turek, F. W. et al. Obesity and metabolic syndrome in circadian Clock mutant mice. *Science* **308**, 1043–1045 (2005).

13. Kohsaka, A. et al. High-fat diet disrupts behavioral and molecular circadian rhythms in mice. *Cell Metab.* **6**, 414–421 (2007).

14. Fei, C. J. et al. Exome sequencing identifies genes associated with sleep-related traits. *Nat. Hum. Behav.* **8**, 576–589 (2024).

15. Jones, S. E. et al. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat. Commun.* **10**, 343 (2019).

16. Soccio, R. E. et al. Genetic variation determines PPARgamma function and anti-diabetic drug response in vivo. *Cell* **162**, 33–44 (2015).

17. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

18. Autieri, M. V. et al. Allograft inflammatory factor-1 expression correlates with cardiac rejection and development of cardiac allograft vasculopathy. *Circulation* **106**, 2218–2223 (2002).

19. Utans, U., Arceci, R. J., Yamashita, Y. & Russell, M. E. Cloning and characterization of allograft inflammatory factor-1: a novel macrophage factor identified in rat cardiac allografts with chronic rejection. *J. Clin. Investig.* **95**, 2954–2962 (1995).

20. Goriki, A. et al. A novel protein, CHRONO, functions as a core component of the mammalian circadian clock. *PLoS Biol.* **12**, e1001839 (2014).

21. Talamanca, L., Gobet, C. & Naef, F. Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in humans. *Science* **379**, 478–483 (2023).

22. Hitsumoto, T. et al. Restoration of cardiac myosin light chain kinase ameliorates systolic dysfunction by reducing superrelaxed myosin. *Circulation* **147**, 1902–1918 (2023).

23. Takahashi, J. S. Transcriptional architecture of the mammalian circadian clock. *Nat. Rev. Genet.* **18**, 164–179 (2017).

24. Mullins, N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet* **53**, 817–829 (2021).

25. Herzog, E. D., Hermanstyne, T., Smyllie, N. J. & Hastings, M. H. Regulating the suprachiasmatic nucleus (SCN) circadian clockwork: interplay between cell-autonomous and circuit-level mechanisms. *Cold Spring Harb. Perspect. Biol.* **9**, a027706 (2017).

26. Fang, B. et al. Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. *Cell* **159**, 1140–1152 (2014).

27. Koritala, B. S. C. et al. Obstructive sleep apnea in a mouse model is associated with tissue-specific transcriptomic changes in circadian rhythmicity and mean 24-hour gene expression. *PLoS Biol.* **21**, e3002139 (2023).

28. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).

29. Koike, N. et al. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* **338**, 349–354 (2012).

30. Tehranchi, A. K. et al. Pooled ChIP-seq links variation in transcription factor binding to complex disease risk. *Cell* **165**, 730–741 (2016).

31. Reinke, H. et al. Differential display of DNA-binding proteins reveals heat-shock factor 1 as a circadian transcription factor. *Genes Dev.* **22**, 331–345 (2008).

32. Besnard, A., Galan-Rodriguez, B., Vanhoutte, P. & Caboche, J. Elk-1 a transcription factor with multiple facets in the brain. *Front. Neurosci.* **5**, 35 (2011).

33. Vanhoutte, P. et al. Opposing roles of Elk-1 and its brain-specific isoform, short Elk-1, in nerve growth factor-induced PC12 differentiation. *J. Biol. Chem.* **276**, 5189–5196 (2001).

34. Farh, K. K. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).

35. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).

36. Corradin, O. & Scacheri, P. C. Enhancer variants: evaluating functions in common disease. *Genome Med.* **6**, 85 (2014).

37. Sollis, E. et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).

38. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era–concepts and misconceptions. *Nat. Rev. Genet.* **9**, 255–266 (2008).

39. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

40. Willer, C. J. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).

41. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).

42. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).

43. Wang, F. et al. LGR4 acts as a link between the peripheral circadian clock and lipid metabolism in liver. *J. Mol. Endocrinol.* **52**, 133–143 (2014).

44. Saponara, E. et al. Loss of hepatic leucine-rich repeat-containing G-protein coupled receptors 4 and 5 promotes nonalcoholic fatty liver disease. *Am. J. Pathol.* **193**, 161–181 (2023).

45. Wang, Z. & Liu, H. Roles of lysine methylation in glucose and lipid metabolism: functions, regulatory mechanisms, and therapeutic implications. *Biomolecules* **14**, 862 (2024).

46. Ye, R., Huang, J., Wang, Z., Chen, Y. & Dong, Y. The role and mechanism of essential selenoproteins for homeostasis. *Antioxidants* **11**, 973 (2022).

47. Zhang, Y. et al. GENE REGULATION. Discrete functions of nuclear receptor Rev-erbalpha couple metabolism to the clock. *Science* **348**, 1488–1492 (2015).

48. Zhou, X. et al. IGF2 deficiency promotes liver aging through mitochondrial dysfunction and upregulated CEBPB signaling in D-galactose-induced aging mice. *Mol. Med.* **29**, 161 (2023).

49. Nicholas, R. S., Stevens, S., Wing, M. G. & Compston, D. A. Microglia-derived IGF-2 prevents TNFalpha induced death of mature oligodendrocytes in vitro. *J. Neuroimmunol.* **124**, 36–44 (2002).

50. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the "Sum of Single Effects" model. *PLoS Genet* **18**, e1010299 (2022).

51. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **82**, 1273–1300 (2020).

52. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

53. Consortium, G. T. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

54. Seney, M. L. et al. Diurnal rhythms in gene expression in the prefrontal cortex in schizophrenia. *Nat. Commun.* **10**, 3355 (2019).

55. Schwartz, P. B. et al. The circadian clock is disrupted in pancreatic cancer. *PLoS Genet.* **19**, e1010770 (2023).

56. Li, L. et al. An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat. Genet.* **53**, 994–1005 (2021).

57. Gate, R. E. et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat. Genet.* **50**, 1140–1150 (2018).

58. Cornelissen, G. Cosinor-based rhythmometry. *Theor. Biol. Med. Model* **11**, 16 (2014).

59. Cheverud, J. M. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* **87**, 52–58 (2001).

60. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221–227 (2005).

61. Davis, J. R. et al. An efficient multiple-testing adjustment for eQTL studies that accounts for linkage disequilibrium between variants. *Am. J. Hum. Genet.* **98**, 216–224 (2016).

62. Weger, B. D. et al. Systematic analysis of differential rhythmic liver gene expression mediated by the circadian clock and feeding rhythms. *Proc. Natl Acad. Sci. USA* **118**, e2015803118 (2021).

63. Thaben, P. F. & Westermark, P. O. Differential rhythmicity: detecting altered rhythmicity in biological data. *Bioinformatics* **32**, 2800–2808 (2016).

64. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283–285 (2016).

65. Perrin, L. et al. Transcriptomic analyses reveal rhythmic and CLOCK-driven pathways in human skeletal muscle. *Elife* **7**, e34114 (2018).

66. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).

67. Hughes, M. E. et al. Guidelines for genome-scale analysis of biological rhythms. *J. Biol. Rhythms* **32**, 380–393 (2017).

68. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinforma.* **14**, 128 (2013).

69. Rozowsky, J. et al. The EN-TEx resource of multi-tissue personal epigenomes & variant-impact models. *Cell* **186**, 1493–1511 e1440 (2023).

70. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

71. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

72. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

73. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

74. Richardson, T. G. et al. Characterising metabolomic signatures of lipid-modifying therapies through drug target Mendelian randomisation. *PLoS Biol.* **20**, e3001547 (2022).

75. Davyson, E. et al. Metabolomic investigation of major depressive disorder identifies a potentially causal association with polyunsaturated fatty acids. *Biol. Psychiatry* **94**, 630–639 (2023).

76. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* **17**, e1009440 (2021).

77. Butte, N. F., Cai, G., Cole, S. A. & Comuzzie, A. G. Viva la Familia Study: genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. *Am. J. Clin. Nutr.* **84**, 673–644 (2006).

78. Comuzzie, A. G. et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One* **7**, e51954 (2012).

79. Sabo, A. et al. Exome sequencing reveals novel genetic loci influencing obesity-related traits in Hispanic children. *Obesity* **25**, 1270–1276 (2017).

80. Uffelmann, E. et al. Genome-wide association studies. *Nat. Rev. Methods Prim.* **1**, 59 (2021).

81. Marees, A. T. et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* **27**, e1608 (2018).

82. Chen, Y. Human genetic variation determines 24-hour rhythmic gene expression and disease risk. https://doi.org/10.5281/zenodo.15127913 (2025).

Author contributions

Y.C. and D.G. conceptualized the study, interpreted data, and wrote the manuscript, which was revised and approved by all authors. Y.C. and P.L. conducted SNP calling based on RNA-seq data. Y.C. and A.S. processed the data from VIVA project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-59524-5.

Correspondence and requests for materials should be addressed to Dongyin Guan.

Peer review information *Nature Communications* thanks Maxime Rotival, Alexandros Simistiras, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.