Universiteit
Leiden
The Netherlands

# Investigating metabolic disease in human induced pluripotent stem cells : apidocyte size, insulin signaling and hepatic lipids

Friesen, M.

**Citation**
Friesen, M. (2018, September 5). *Investigating metabolic disease in human induced pluripotent stem cells : apidocyte size, insulin signaling and hepatic lipids*. Retrieved from https://hdl.handle.net/1887/64936

| | |
|---|---|
| Version: | Not Applicable (or Unknown) |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/64936 |

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden


The handle http://hdl.handle.net/1887/64936 holds various files of this Leiden University dissertation.

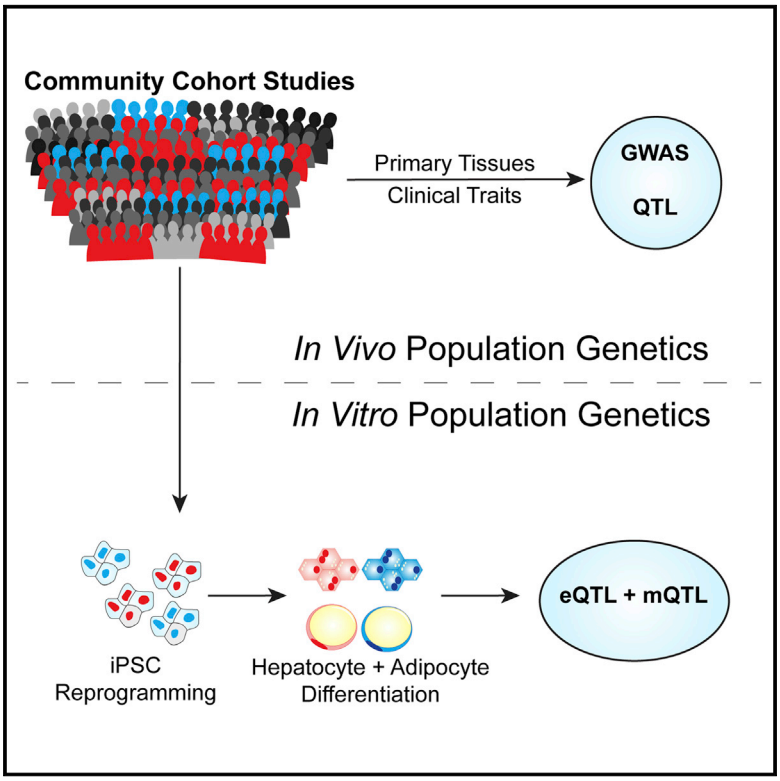
**Author**: Friesen, M.
**Title**: Investigating metabolic disease in human induced pluripotent stem cells : apidocyte size, insulin signaling and hepatic lipids
**Issue Date**: 2018-09-05

Resource

Cell Stem Cell

# Induced Pluripotent Stem Cell Differentiation Enables Functional Validation of GWAS Variants in Metabolic Disease

## Graphical Abstract




## Authors

Curtis R. Warren, John F. O'Sullivan, Max Friesen, ..., Ramachandran S. Vasan, Christopher J. O'Donnell, Chad A. Cowan

## Correspondence

chad_cowan@harvard.edu

## In Brief

As part of the NHLBI NextGen consortium, Warren et al. show that analyzing the cellular pathophysiology of cells differentiated from multiple lines of an iPSC library is a promising complementary approach in the functional analysis of GWAS variants.

## Highlights

- Peripheral blood cells from study participants were reprogrammed to iPSCs
- iPSC lines differentiated into white adipocytes and hepatocyte-like cells
- Transcriptomic and metabolomic profiling highlights cellular phenotypes
- Differentiated cells reproduce expression of variation-associated loci

Warren et al., 2017, Cell Stem Cell *20*, 547–557
April 6, 2017 © 2017 Elsevier Inc.
http://dx.doi.org/10.1016/j.stem.2017.01.010

Cell Stem Cell
Resource

# Induced Pluripotent Stem Cell Differentiation Enables Functional Validation of GWAS Variants in Metabolic Disease

Curtis R. Warren,[1] John F. O'Sullivan,[2] Max Friesen,[1] Caroline E. Becker,[1] Xiaoling Zhang,[8,10] Poching Liu,[3] Yoshiyuki Wakabayashi,[3] Jordan E. Morningstar,[2] Xu Shi,[2] Jihoon Choi,[1] Fang Xia,[1] Derek T. Peters,[1] Mary H.C. Florido,[1] Alexander M. Tsankov,[1,4] Eilene Duberow,[1] Lauren Comisar,[1] Jennifer Shay,[1] Xin Jiang,[1] Alexander Meissner,[1,4] Kiran Musunuru,[1] Sekar Kathiresan,[5] Laurence Daheron,[1] Jun Zhu,[3] Robert E. Gerszten,[2] Rahul C. Deo,[6] Ramachandran S. Vasan,[7,8,9] Christopher J. O'Donnell,[10,11] and Chad A. Cowan[1,4,12,13,*]
[1]Department of Stem Cell and Regenerative Biology and Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138, USA
[2]Cardiovascular Research Center and Cardiology Division, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[3]DNA Sequencing and Genomics Core, National Heart Lung and Blood Institute, NIH, Bethesda, MD 20892, USA
[4]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[5]Center for Human Genetic Research and Cardiovascular Research Center, Massachusetts General Hospital, 185 Cambridge Street, CPZN 5.252, Boston, MA 02114, USA
[6]Cardiovascular Research Institute, Department of Medicine and Institute for Human Genetics, University of California, San Francisco, and California Institute for Quantitative Biosciences, San Francisco, CA 94143, USA
[7]The Framingham Heart Study, Sections of Preventive Medicine and Epidemiology and Cardiology, Framingham, MA 01702, USA
[8]School of Medicine, Boston University, Boston, MA 02118, USA
[9]School of Public Health, Boston University, Boston, MA 02118, USA
[10]The Framingham Heart Study, Population Sciences Branch, Division of Intramural Research, National Heart, Lung and Blood Institute, Framingham, MA 01702, USA
[11]Cardiology Section, Department of Medicine, Boston Veterans Administration Healthcare and Brigham and Women's Hospital, Boston, MA 02114, USA
[12]Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[13]Lead Contact
*Correspondence: chad_cowan@harvard.edu
http://dx.doi.org/10.1016/j.stem.2017.01.010

## SUMMARY

**Genome-wide association studies (GWAS) have highlighted a large number of genetic variants with potential disease association, but functional analysis remains a challenge. Here we describe an approach to functionally validate identified variants through differentiation of induced pluripotent stem cells (iPSCs) to study cellular pathophysiology. We collected peripheral blood cells from Framingham Heart Study participants and reprogrammed them to iPSCs. We then differentiated 68 iPSC lines into hepatocytes and adipocytes to investigate the effect of the 1p13 rs12740374 variant on cardiometabolic disease phenotypes via transcriptomics and metabolomic signatures. We observed a clear association between rs12740374 and lipid accumulation and gene expression in differentiated hepatocytes, in particular, expression of *SORT1*, *CELSR2*, and *PSRC1*, consistent with previous analyses of this variant using other approaches. Initial investigation of additional SNPs also highlighted correlations with gene expression. These findings suggest that iPSC-based population studies hold promise as tools for the functional validation of GWAS variants.**

## INTRODUCTION

Genome-wide association studies (GWASs) provide insight into the effects of common genetic variants on physiology and disease. By coupling this technological breakthrough with the clinical resource of large community-based longitudinal studies, GWASs have revealed thousands of novel genetic mechanisms underlying disease and other traits (Teslovich et al., 2010; Visscher et al., 2012). In addition to associating coding mutations and novel genes with disease, GWASs have identified subtle phenotypes driven by non-coding SNPs by virtue of their large sample size and unbiased study design. Hundreds of non-coding SNPs have been implicated by GWASs in the control of blood lipid concentrations and/or myocardial infarction (MI) that results from dyslipidemia. One of the loci most significantly associated with variance in plasma concentration of low-density lipoprotein cholesterol (LDL-C) and MI is located on chromosome 1p13. An expression quantitative trait locus (eQTL) analysis of non-coding SNPs at 1p13 correlated the minor haplotype with increased gene expression specifically in human liver (Musunuru et al., 2010). One SNP, rs12740374, was responsible for the minor haplotype-associated increase in gene expression of *SORT1*, *PSRC1*, and *CELSR2*. The minor allele of rs12740374 mediates this effect by creating a CAAT-enhancer-binding protein (C/EBP) recognition site (Musunuru et al., 2010). We found that increasing hepatic *SORT1* expression resulted in decreased secretion of LDL-C, explaining the decreased risk of MI

associated with this common non-coding variant. This finding prompted us to explore whether induced pluripotent stem cells (iPSCs) could be used to model the effect of this non-coding SNP on hepatic lipid phenotypes. In addition to employing genome editing to delete the causal SNP in an iPSC line, we have also, for the first time, performed a hepatic eQTL analysis using a large cohort of iPSCs from Framingham Heart Study (FHS) participants of variable genetic backgrounds. By employing rigorous reprogramming and differentiation efficiency controls as well as cell purification strategies, we have demonstrated that iPSCs are indeed capable of recreating a hepatocyte-specific eQTL and, in so doing, have garnered insights into hepatic lipid biology driven by the common non-coding variant rs12740374.

iPSCs can be directly derived from the somatic cells of any donor prospectively identified to carry a genotype of interest (Park et al., 2008). This versatile feature makes iPSCs an ideal platform for the study of Mendelian or monogenic disease. The use of iPSCs for the study of common SNPs with subtle effects on metabolic phenotypes may require large-scale reprogramming efforts coupled with a fine control of directed differentiation techniques. As a proof of principle for future "GWAS in a dish" approaches to common disease genetics, we chose to model the eQTL previously characterized at the 1p13 locus that is associated with increased *SORT1* expression. In vivo, an increase in *SORT1* expression occurs specifically in the liver tissue but not in the adipose depots. This modeling experiment involved the reprogramming of somatic cells from 34 donor participants selected for their homozygous major or minor genotype at the 1p13 SNP rs12740374 and differentiation of the cognate iPSCs into both hepatocytes and white adipocytes. RNA sequencing, metabolomics, and integrative statistical analyses were employed to analyze the effect of the 1p13 genotype on gene expression and metabolism. This experimental approach yielded molecular insights into the protective mechanism of the 1p13 SNP, identifying dysregulation of lipid accumulation and a number of novel *trans*-eQTLs specifically in differentiated hepatocytes. These results demonstrate the capacity for iPSCs to recapitulate expression quantitative trait locus analyses in vitro and yield novel pathophysiological insights. With improved reprogramming and differentiation protocols, this strategy may be used for discovery and functional studies of common disease-related genetic variants in the future.

## RESULTS

### Recruitment of the Donor Cohort and Reprogramming of Peripheral Blood Mononuclear Cells

The FHS offspring cohort, consisting of offspring and spouses of offspring of the original FHS cohort, was enrolled in 1971 as a prospective study of heart disease epidemiology (Kannel et al., 1979). This cohort of 5,124 individuals of European descent has been serially examined at regular intervals (Kannel et al., 1979). The offspring database now includes a large number of biomarker measurements in parallel with subclinical and clinical phenotypes as well as genome-wide genotyping and genome sequencing data measured in many participants. We collected T cells from consenting offspring cohort participants in the ninth examination cycle (2011–2014). We reprogrammed cryopreserved peripheral blood mononuclear cells (PMBCs) from a selected subset of 17 participants homozygous for the major haplotype at the 1p13 locus and 17 participants homozygous for the minor haplotype at the 1p13 locus (Figure 1A). The non-replicating Sendai virus (SeV) was used to introduce the reprogramming factors OCT4, KLF4, SOX2, and c-Myc into T cells (Takahashi et al., 2007). Five to seven clonal cell lines were isolated from the reprogrammed cells for each donor participant, and two were further expanded and characterized, resulting in 68 iPSC lines.

Each iPSC line was submitted to a battery of tests for quality control purposes. The expression of key pluripotency markers (OCT4, Nanog, SSEA3, SSEA4, and Tra-1-60) was assessed using immunofluorescence (Figure 1B). qRT-PCR was used to assess the expression of six key pluripotency genes (*DNMT3B*, *TERT*, *NANOG*, *OCT4*, *REX1*, and *SOX2*). The expression of each of these pluripotency genes was detected in all 68 iPSC lines and was equivalent to the expression of these pluripotency genes by an existing panel of human embryonic stem cell (hESC) lines (Figure 1C). G-band karyotyping was performed to ensure that these cell lines had not incurred chromosomal abnormalities (Figure S1A). Sixty-four of the 68 iPSC lines carried no chromosomal abnormalities (Table S1). Two cell lines carried three X chromosomes, and two cell lines carried a non-pathological inversion on chromosome 11. In both cases, these abnormalities occurred in clonal cell lines derived from the same participant, suggesting that these chromosomal abnormalities stem from the donor participant's genome rather than the reprogramming process. These chromosomal abnormalities had no effect on differentiation efficiency (Table S1) because efficiencies of these cell lines were within 1 SD of the mean. Similarly, expression of adipocyte and hepatocyte-like cell (HLC) marker genes was not affected in these cell lines. A common concern of reprogramming is the persistence of transgenic reprogramming factors after pluripotency is achieved. The delivery method we used relied on the non-integrating SeV, which is diluted out of the cellular population through passaging. We assessed the presence of SeV-delivered transgenes via qRT-PCR at varying passages and, ultimately, could not detect the persistence of SeV RNA in our cell lines (Figure S1B). All cell lines were characterized by short tandem repeat (STR) fingerprinting, and all were deemed matched to their donor of origin by this analysis (data not shown).

We employed the embryoid body (EB) Scorecard assay (Bock et al., 2011; Tsankov et al., 2015) as a measure of iPSC differentiation potential prior to differentiation into functional cell types. After differentiation into EBs and plating onto gelatin, RNA was harvested, and the expression of 92 genes was assayed by the Taqman-based Scorecard assay. These genes were chosen for their specificity in expression to the three germ layers and pluripotent stem cells. EBs from all cell lines lost expression of pluripotency genes and expressed markers for ectoderm, mesoderm, and endoderm (Figure 1D). Some variability was evident between individual iPSC lines in their capacity to differentiate into the three germ layers (Figure S1C), but every cell line expressed genes associated with each germ layer during EB differentiation, confirming their pluripotent state.

### Programming of iPSCs into White Adipocytes

We differentiated white adipocytes from each of our 68 iPSC lines. To control for variability in differentiation, we performed

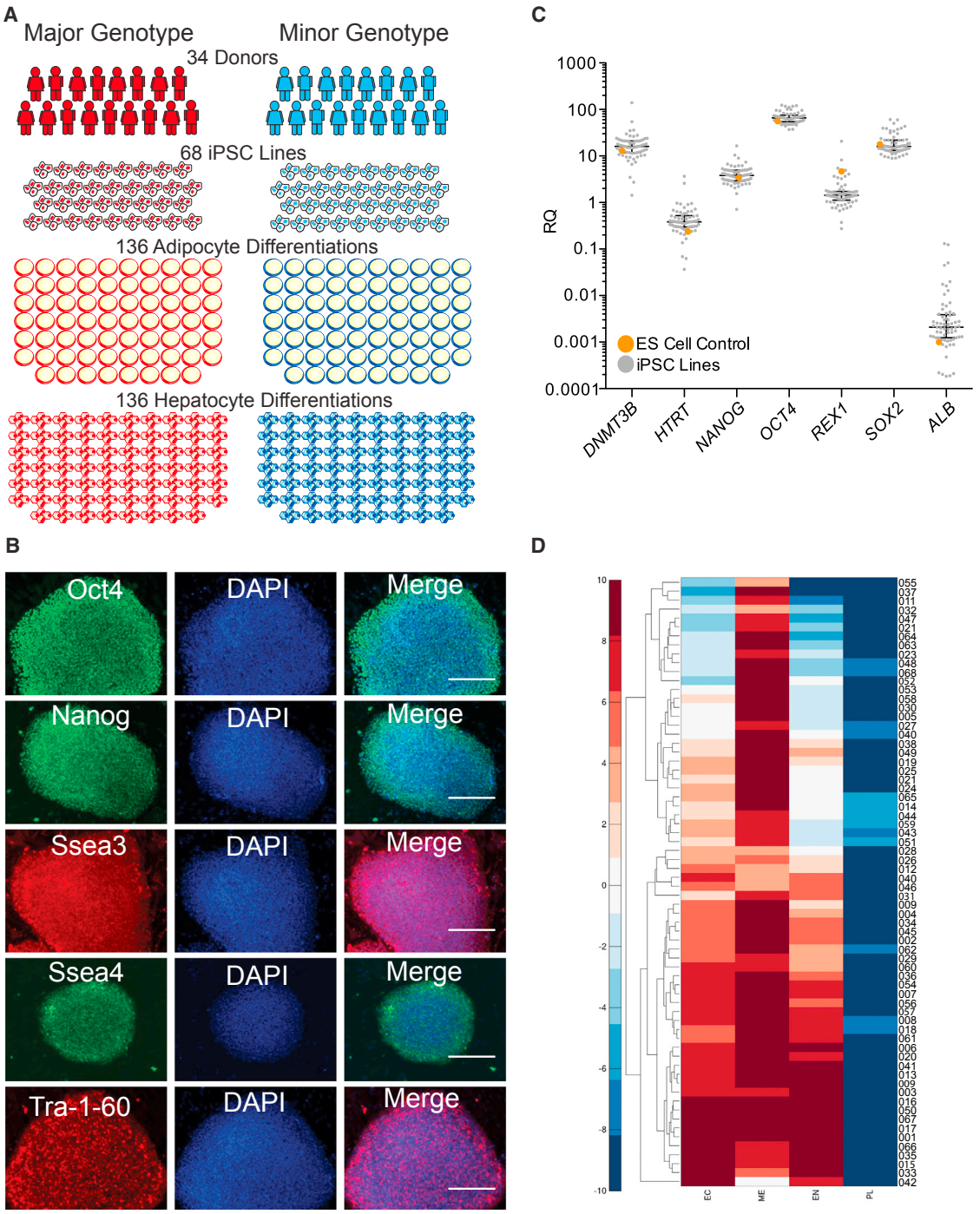



**Figure 1. Research Strategy and Reprogramming of 34 Donor Participant Samples into iPSCs**

(A) Blood samples from 17 donors of the major genotype and 17 donors of the minor genotype for rs12740374 were reprogrammed into two clonal iPSC lines each. These cell lines were then differentiated into both white adipocytes and HLCs for metabolic and transcriptomic profiling.

(B) Representative immunofluorescence indicating expression of pluripotency markers by iPSC colonies. Scale bars, 200 μm.

(C) qRT-PCR was performed to demonstrate expression of pluripotency genes by each cell line (*ALB* served as a negative control) (n = 68). Orange dots, human embryonic stem cell line control value; line, mean; bars, SEM. Relative quantification (RQ) as described in the STAR Methods.

(D) Hierarchical clustering of cell lines (rows) by Euclidean distance similarity of their Scorecard differentiation potential (columns). The Scorecard differentiation potential was computed as in Tsankov et al. (2015) using gene expression signatures of ectoderm (EC), mesoderm (ME), endoderm (EN), and pluripotent (PL) marker genes. The legend on the left displays the range of scores relative to a reference set of 23 human PSC lines. Clone IDs are listed on the right y axis. See also Figure S1.

two experimental replicates for each cell line, resulting in 136 total replicates of white adipocyte differentiation (Figure 1A). We followed a published white adipocyte programming protocol (Ahfeldt et al., 2012; Lee and Cowan, 2014), relying on differentiation into mesenchymal progenitor cells (MPCs) followed by programming into white adipocytes by a lentivirally delivered and doxycycline-inducible peroxisome proliferator activated receptor gamma-2 (PPARG2) (Figure 2A).

MPCs were produced in batches, cryopreserved, and RNA-profiled prior to white adipocyte differentiation. MPC expression of key mesoderm marker genes (*CD44*, *NT5E*, and *ENG*) was confirmed by qRT-PCR (Figure S2B). Two MPC differentiations were performed for each clonal cell line, and the MPC batch with the highest steady-state expression of these three mesoderm genes was chosen for programming with PPARG2. PPARG2 was induced by doxycycline for 16 days, and, 6 days after withdrawal of doxycycline, differentiation efficiency was assessed by immunofluorescence (Figure 2B). C/EBP-alpha works in concert with PPARG2 to promote adipogenesis and establish white adipocyte maturity, and its presence is used here as a proxy for differentiation efficiency (Freytag et al., 1994; Wu et al., 1999). A high-content automated imaging platform was used to analyze the cells in 25 fields of a representative well of differentiated white adipocytes (Figure S2A). Differentiation was quantified for each replicate (Figure 2C). The mean cumulative differentiation efficiency over all 136 differentiation experiments was 60.4% and the median differentiation efficiency over these experiments was 67.7%.

RNA, triglycerides, and metabolites were collected from each batch of differentiated cells. Expression of *ADIPOQ* and accumulation of triglycerides correlated with differentiation efficiency (Figures 2D and 2E). Although the differentiation efficiency of some cell lines was below the mean value of 67%, all cell lines expressed *FABP4* following doxycycline treatment (Figure S2D). *FABP4* is a direct transcriptional target of PPARG2, indicating viral transduction and activity of transgenic PPARG2 even in poorly differentiated cell lines.

## Directed Differentiation of iPSCs into Hepatocyte-like Cells

We used fluorescence-activated cell sorting (FACS) to prospectively isolate HLCs derived by a directed differentiation protocol (Peters et al., 2016). As with the differentiation of white adipocytes, each iPSC line was differentiated twice to account for variability, resulting in 136 HLC differentiations (Figure 1A). Using asialoglycoprotein receptor 1 (ASGR1) as a surface marker of hepatocytes, we purified HLCs from the heterogeneous population of cells produced by differentiation through definitive endoderm (DE), hepatic endoderm (HE), immature hepatocytes (IMHs), and, finally, into mature hepatocytes (MHs) (Figure 2F). HLCs express albumin and ASGR1 (Figure 2G) and perform lipid processing in a biologically relevant fashion responsive to perturbations of sortilin expression (Ding et al., 2013). Differentiation efficiencies for each of the 68 iPSC lines was measured by ASGR1 FACS (Figure 2H) and varied from 65% to 0.6%. The mean and median differentiation efficiencies were 13% and 9.9%, respectively.

Expression of the hepatocyte markers *ALB* and *APOB* by FACS-isolated HLCs was analyzed by qRT-PCR (Figures 2I and 2J). Nearly all isolated HLC populations expressed *ALB* and *APOB*, and the expression levels of these two genes were positively correlated (Pearson correlation coefficient [r] = 0.48). In a minority of cell lines where differentiation efficiency was very low, the expression of *ALB* and *APOB* was similarly diminished, equivalent to expression of the negative control gene *CD4*. Sorting of non-specific cells was diminished by the use of an isotype control antibody for each FACS replicate (Figure S2E). For modeling of the lipid phenotype associated with the 1p13 genotype of these cells, conditioned medium and metabolites were extracted from adherent unsorted HLCs.

## Modeling of the eQTL Occurring at 1p13 Accurately Recreates the In Vivo Phenotype

The minor haplotype at the 1p13 locus has been associated with decreased plasma LDL-C concentrations and a decreased risk of MI, mediated by increased hepatic sortilin expression (Musunuru et al., 2010). Three genes (*CELSR2*, *PSRC1*, and *SORT1*) are upregulated in hepatocytes through creation of a putative C/EBP transcription factor binding site by the causal SNP (rs12740374; Figure 3A). In vitro modeling of the hepatic eQTL at this locus accurately reproduced the previously published expression profile from liver biopsies, with *CELSR2*, *PSRC1*, and *SORT1* increased in ASGR1-positive HLCs carrying the minor 1p13 locus haplotype compared with matched cells carrying the major 1p13 locus haplotype (Figure 3B). Although significant ($p = 0.04$), the increase in *SORT1* expression is modest (1.3-fold). Expression of *CELSR2*, *PSRC1*, and *SORT1* does not appear to correlate with differentiation efficiency (Figures S3D–S3F) because FACS isolation of ASGR1-positive HLCs eliminates nearly all undifferentiated cells (isotype controls are used to gate out undifferentiated cells). In contrast, white adipocytes demonstrated no change in *SORT1* expression between the major and minor haplotype (Figure 3D), consistent with previously published data from primary adipose tissue samples (Figure 3D), and *SORT1* expression did not correlate with differentiation efficiency in white adipocytes (Figure S3G).

To validate our experimental findings, we used clustered regularly interspaced palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) genome editing techniques to knock out the causal SNP rs12740374 in a cell line carrying two minor alleles at the 1p13 locus (Figure S3A). The resultant knockout cell lines carried deletions of rs12740374 and surrounding base pairs as their only genetic difference from the isogenic parental and wild-type cell lines. Each mutant cell line incurred homozygous deletion of the rs12740374 base pair with heterozygous deletions of the surrounding base pairs. The largest deletion included 27 base pairs and the entire putative C/EBP transcription factor binding site, and the smallest deletion removed three base pairs and inserted three base pairs, leaving four base pairs of the putative C/EBP transcription factor binding site. In each case, the rs12740374 base pair was deleted, removing the critical thymine in the canonical C/EBP transcription factor binding site created by the minor allele of this SNP. We differentiated these four knockout clones and two wild-type clones into HLCs and isolated ASGR1-positive cells by FACS. In this experiment, we observed a 1.3-fold increase in expression of *SORT1* in ASGR1-positive HLCs derived

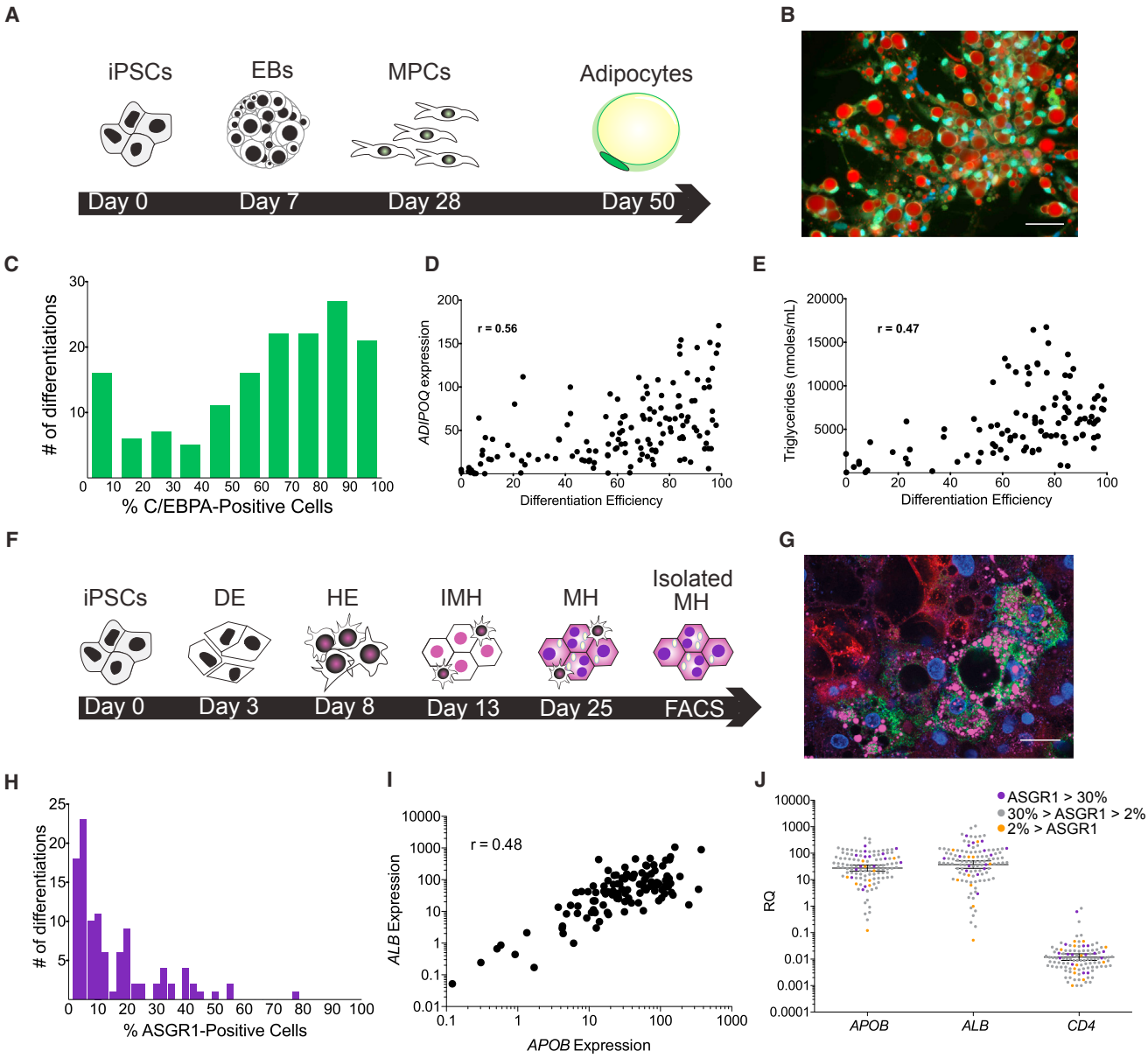



**Figure 2. Differentiation of iPSCs into Adipocytes and HLCs**

(A) Differentiation protocol for production of white adipocytes.

(B) Immunofluorescence indicating expression of C/EBP-alpha (green) and accumulation of triglycerides (red). Nuclei, blue. Scale bar, 100 μm.

(C) Differentiation efficiencies recorded for each differentiation experiment.

(D and E) Expression of *ADIPOQ* correlates with differentiation efficiency (D, n = 136), as does the accumulation of triglycerides (E, n = 90).

(F) Directed differentiation protocol for production and isolation of HLCs.

(G) Immunofluorescence indicating expression of the hepatocyte markers ASGR1 (green) and albumin (red) as well as accumulation of triglycerides (pink). Hoechst, blue. Scale bar, 50 μm.

(H) Differentiation efficiencies for each iPSC line as measured by flow cytometry for surface ASGR1 expression (n = 136).

(I) Correlation of *ALB* and *APOB* expression in ASGR1-positive HLCs (n = 136).

(J) qRT-PCR data indicating expression of *APOB* and *ALB* and non-existent expression of *CD4* (as a negative control) (n = 136). Line, mean; bars, SEM. RQ as described in the STAR Methods.

See also Figure S2.

from the minor haplotype parental cells compared with the genome-edited clonal cell lines (Figure 3C). This finding matched precisely that of the in vitro eQTL that we demonstrated in the cohort of iPSC-derived hepatocytes (Figure 3B). Both the genome-edited clonal cell lines and parental cell lines demonstrated more than 10% differentiation efficiency prior to FACS (Figure S3B), and there was no difference in *ALB* expression between these cell lines after sorting (Figure S3C).

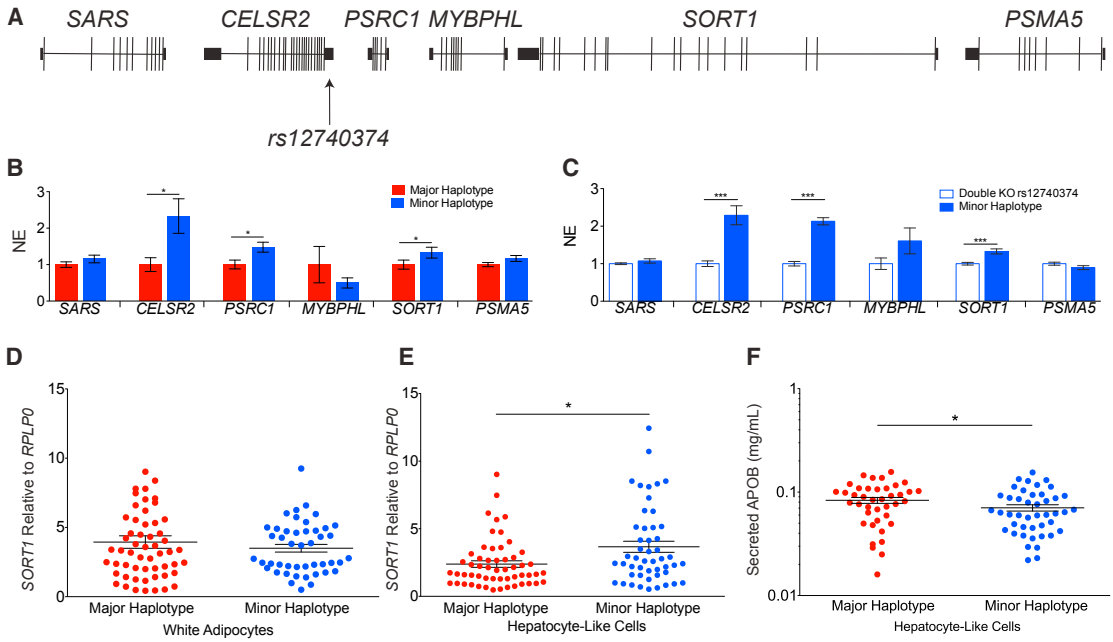



**Figure 3. Phenotyping of Differentiated HLCs and Adipocytes Demonstrates In Vitro Recapitulation of the 1p13 *SORT1* eQTL**

(A) The 1p13 locus encodes six genes in linkage disequilibrium, and the causal SNP (rs12740374) associated with decreased serum LDL cholesterol and increased expression of three genes is located in the 3′ UTR of *CELSR2*.

(B) qRT-PCR analysis of ASGR1-positive HLCs homozygous for the major and minor haplotypes at 1p13 (n = 68). NE, normalized expression, normalized to major genotype. *CELSR2* p = 0.016, *PSRC1* p = 0.011, *SORT1* p = 0.046, Student's t test.

(C) qRT-PCR analysis of ASGR1-positive genome-edited HLCs (n = 18 wild-type, n = 20 mutant). NE, normalized expression, normalized to minor genotype. *CELSR2*, *PSRC1*, and *SORT1* p values < 0.001, unpaired t test.

(D) Scatterplot of qRT-PCR for *SORT1* in differentiated adipocytes homozygous for the major or minor haplotype at the 1p13 locus (n = 68).

(E) Scatterplot of qRT-PCR for *SORT1* in ASGR1-positive HLCs homozygous for the major or minor haplotype at the 1p13 locus (n = 68). p = 0.019, Student's t test.

(F) Scatterplot of APOB accumulation in HLC-conditioned medium as measured by ELISA. p = 0.0258, Student's t test.

Line in scatterplots, mean; bars in all plots, SEM. See also Figure S3.

Apolipoprotein B (APOB) secretion has been used as an accurate surrogate of LDL secretion in vitro. We collected conditioned medium from differentiated cells and performed an ELISA to measure the effect of 1p13 genotype on HLC secretion of APOB. HLCs carrying the minor genotype secreted significantly (30%) less APOB than major genotype HLCs (Figure 3F). This is consistent with previous findings from human serum (Musunuru et al., 2010). APOB secretion correlated with differentiation efficiencies for each cell line (Figure S3H), and there was no statistical difference in HLC differentiation efficiency between cells carrying the major or minor genotypes (Figure S3I).

## Cell Transcriptomes Accurately Recreate In Vivo Phenotypes and the 1p13 eQTL

RNA sequencing analyses were performed on samples from iPSCs, white adipocytes, and ASGR1-positive HLCs from all 68 cell lines. Cellular identity and *SORT1* expression related to the 1p13 genotype were the focus of the preliminary analysis of these datasets. The transcriptional profile of each cell type was distinct, as indicated by transcriptome-wide heatmaps (Figure S4A). Tissue-specific biological processes were upregulated in each cell type, as evidenced by gene ontology analysis (Figures S4B–S4D). iPSCs, white adipocytes, and HLCs all cluster separately in a principal-component analysis (PCA) analysis of their gene expression profile (Figure S4E), suggesting that differentiation produced transcriptionally distinct cell types. Cell type-specific gene expression results obtained by qRT-PCR were also replicated in the RNA sequencing dataset (Figure S4F). The gene expression profiles of clonal cell lines derived from the same donor correlated tightly (Figure S4G).

1p13 *SORT1* eQTL modeling was replicated in the RNA sequencing (RNA-seq) dataset by applying a linear mixed model analysis using data from iPSCs, white adipocytes, and HLCs. The predicted eQTL was replicated accurately in all five genes of the 1p13 locus, with *PSRC1*, *CELSR2*, and *SORT1* each being statistically upregulated in minor genotyped HLCs (Figure 4A). *PSRC1* was also upregulated in minor genotyped adipocytes and iPSCs. As expected, there was no expression phenotype at either *SARS* or *PSMA5* in any cell type. This phenotype was also corroborated when analyzing the number of fragments per kilobase per million reads (FPKM), an alternative measure of transcript abundance (Figure 4B). This result is evidence that iPSCs can be used through directed differentiation to model phenotypes driven by common, non-coding variants.

In addition to the chromosome 1p13 eQTL findings presented in Figure 4A, a number of novel *trans*-eQTLs were identified in HLCs (Table S2). Included are the genes that were below the false discovery rate of 10%. No gene ontology process terms

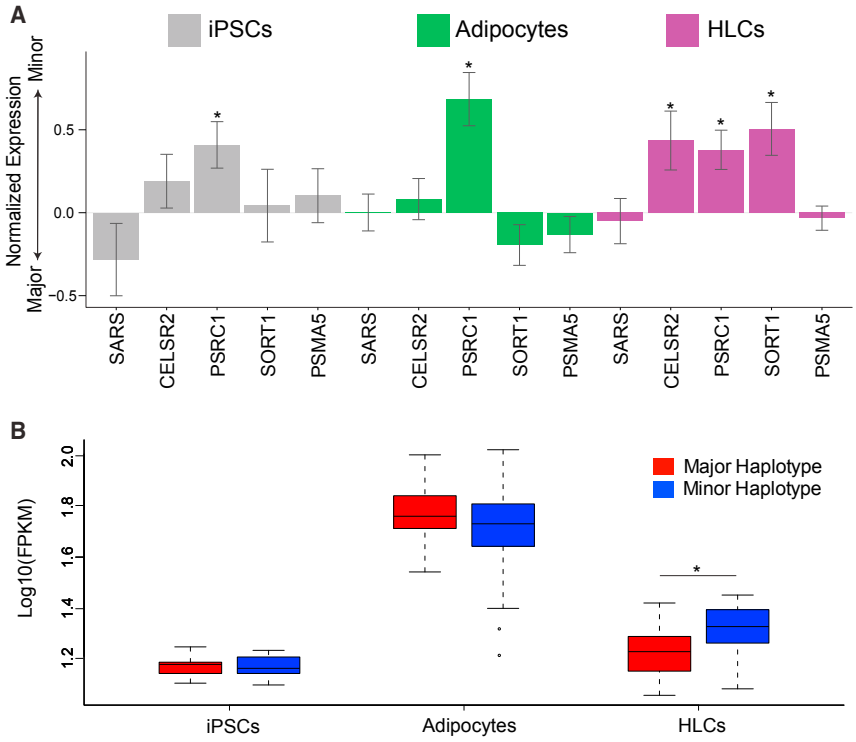



**Figure 4. RNA Sequencing Analyses Support Accurate Modeling of the 1p13 *SORT1* eQTL**

(A) RNA sequencing data comparing expression of genes in the chromosome 1p13 linkage disequilibrium block in cells carrying the major or minor haplotype. Values above and below the x axis display increased expression in cells carrying the minor or major haplotype, respectively (n = 68). FDR values are summarized in Table S3.

(B) *SORT1* expression in iPSCs, adipocytes, and HLCs is measured as FPKM. Bars, mean; whiskers, SD. Outliers in adipocyte data are plotted as dots. p Value from HLCs = 0.034, Student's t test.

See also Figure S4.

were enriched in this gene set. There were no *trans*-eQTLs identified in either iPSCs or adipocytes.

Encouraged by the clear association of the 1p13 genotype with *SORT1* expression in our dataset, we replicated additional eQTL analyses (Table S4). 50 loci previously identified by eQTL analysis in liver tissue (Schadt et al., 2008; Teslovich et al., 2010) were tested for their effect on gene expression. Two SNPs (rs2929282 and rs3177928) correlated with expression of FRMD5 and HLA-DRB1 with false discovery rates (FDRs) of 9.5% and 3.3%, respectively. rs10889353 and rs10195252 correlated with expression of DOCK7 and GRB14, respectively (both at an FDR of 11.6%). The other 46 SNPs were not significantly correlated with gene expression in HLCs.

## Metabolomics Analyses Indicate a General Downregulation of Lipid Accumulation in Minor Genotyped HLCs

Intracellular metabolite extraction was performed as described previously by our group (Patsch et al., 2015). Metabolomic profiling was performed on all three cell types using two targeted profiling methods. The first method was run in positive ion mode and concentrated on lipid subspecies. The second method, the Amide method, was run in negative ion mode and focused on nucleotide and nucleoside phosphates, high-energy intermediates, organic acids, tricarboxylic acid (TCA) cycle intermediates, and glycolytic intermediates. Both methods indicate distinct metabolic profiles for each of the three cell types, as indicated by PCA plots (Figures S5A and S5B).

In both iPSCs (Figure S5C) and adipocytes (Figure S5D), there were no significant differences in lipid metabolites between the minor and major genotype. In iPSCs, the Amide method revealed no significant differences between the major and minor genotype (Figure S5E). In adipocytes, there was a significant decrease in the levels of the nucleoside cytidine diphosphate (CDP) (FDR < 5%) in the minor genotype and a significant increase in the levels of the ubiquitous organic aminosulfonic acid taurine (FDR < 5%) in the minor genotype (both are indicated in green in Figure S5F).

In the HLCs, there were many metabolites with significant differences between minor and major genotypes (indicated in purple in Figures 5A and 5B). The greatest differences were seen in the lipid metabolites, with some species enriched up to 4-fold in major genotype HLCs versus minor genotype cells. The minor genotype HLCs accumulated significantly less phosphatidylethanolamine, triacylglycerol, diacylglycerol, phosphatidylcholine, and sphingomyelin species (FDR < 5%; Figure 5B, indicated in blue), suggesting a non-specific downregulation of lipid synthesis in the minor genotype HLCs. Each bar in Figure 5B represents a subspecies of lipid metabolite; e.g., PC 36:2, where 36 is the number of carbons and 2 is the number of double bonds. Genotype-driven lipid dysregulation occurs across a wide range of subspecies within each lipid class. Minor genotype HLCs also retained significantly less acetoacetic acid (FDR < 5%; Figure 5C), a ketone body whose main function is to provide acetoacetyl coenzyme A (CoA) and acetyl CoA for synthesis of cholesterol, fatty acids, and complex lipids. 2-Aminoadipic acid, a metabolite known to be associated with cardiometabolic disease (Wang et al., 2013), was also significantly decreased in minor genotype HLCs (FDR < 5%; Figure 5C), as were D-gluconic acid, kynurenine, and pseudouridine (FDR < 5%; Figure 5C).

## DISCUSSION

Community-based cohorts have increased the power of human genetics, rendering possible the identification of many novel disease loci. GWAS and eQTL analyses overcame the limitations of traditional human genetics by coupling very large participant cohorts with relatively inexpensive whole-genome genotyping chips and clinical data regarding markers of disease. In this way, subtle effects of common variants were extracted from the noise of genome-wide variation (Teslovich et al., 2010).

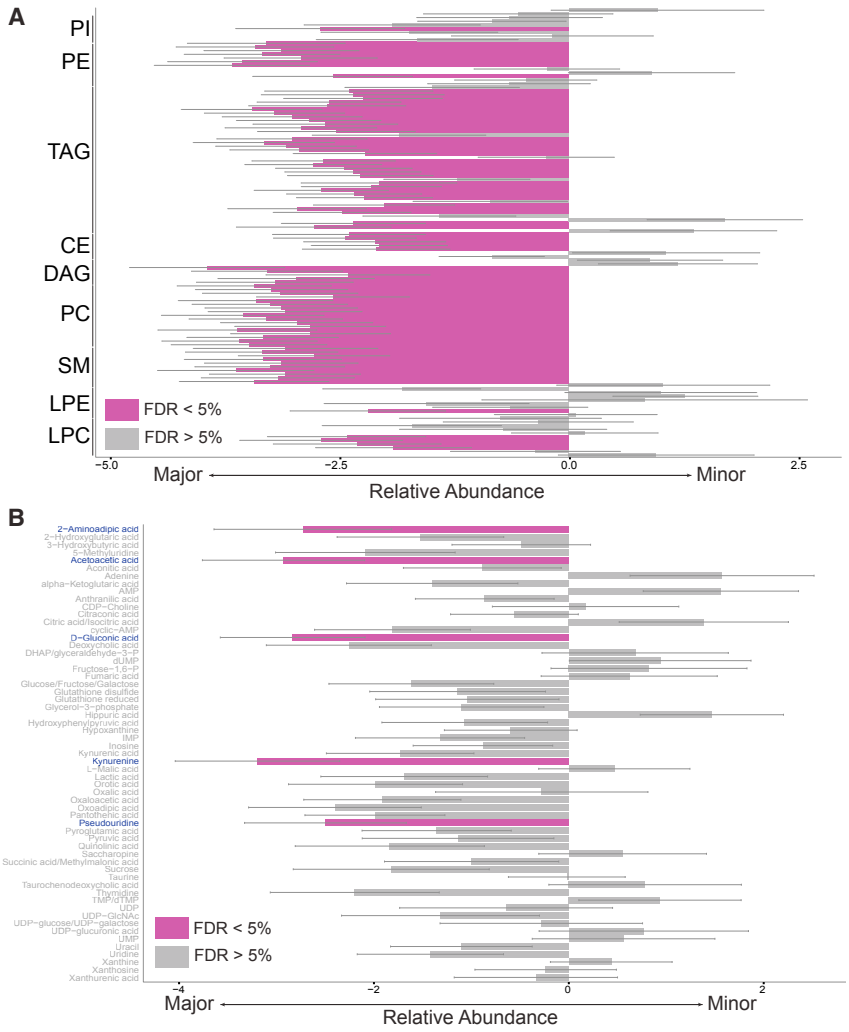



**Figure 5. Metabolomic QTL Findings in Differentiated HLCs**

(A and B) Lipid (A) and Amide metabolomics (B) profiling of HLCs. Analyte classes are indicated on the y axis. For legibility, only the lipid classes are labeled in the y axis, and each bar represents a subspecies of lipid within the larger lipid class. Subspecies are labeled in full in parallel charts Figures S5C and S5D. Bars to the left or right of the y axis indicate enrichment in cells carrying the major or minor haplotype at chromosome 1p13, respectively. Values that satisfy the 5% FDR rate are highlighted in purple. FDR values are summarized in Table S3. PI, phosphatidylinositols; PE, phosphatidylethanolamines; TAG, triacylglycerols; CE, cholesteryl ester; DAG, diacylglycerols; PC, phosphatidylcholines; SM, sphingomyelins; LPC, lysophosphatidylcholines; LPE, lysophosphatidylethanolamines. n = 408.
See also Figure S5.

Simultaneously, the advent of iPSC technology increased the utility of cellular disease models for the study of genotype-phenotype interactions (Takahashi et al., 2007). The capacity to generate pluripotent cells from the somatic cells of any participant or patient of interest provides a direct insight into the relationship between that individual's genotype and the functional capacity of the differentiated cells (Park et al., 2008). Here, for the first time, we have demonstrated that iPSC technology may be feasibly applied to in vitro genetics, given the appropriate cohort size and a carefully planned experiment. This study design eliminates a limitation of the traditional GWAS or eQTL analysis: availability of differentiated cells for mechanistic study of novel disease-associated loci. The advantages of such future in vitro population genetics studies would be manifold. Most notably, future in vitro studies would make possible the correlation of cellular and subcellular phenotypes with genome-wide variation, especially in cell types that are difficult to procure in large quantities from primary tissues, such as neurons or β cells.

We have modeled the association between rs12740374 and hepatic expression of *SORT1*, previously reported to protect from MI secondary to decreased circulating LDL-C concentrations (Musunuru et al., 2010). iPSCs derived from participants ascertained at the rs12740374 locus were differentiated into HLCs and accurately recapitulated gene expression profiles in genes centered on the 1p13 locus. For future in vitro eQTL analyses studying hepatic eQTLs, the *SORT1* phenotype will be useful as a positive control, given that the cohort is large enough and includes participants homozygous for the major and minor haplotypes at chromosome 1p13. iPSCs and cognate adipocytes demonstrated no effect of genotype on expression of *SORT1*, in keeping with in vivo findings.

It is important to note that the effect size of genotype on hepatic gene expression at 1P13, as measured by qRT-PCR, is smaller than that measured previously measured in primary tissues by qRT-PCR and cDNA microarray in primary tissues. It is possible that this discrepancy in effect size is due to the relative "immaturity" of HLCs in comparison with primary hepatocytes (Si-Tayeb et al., 2010) or due to heterogeneity in differentiation. Differentiation of the iPSC cohort and generation of isogenic genome-edited knockout clones of rs12740374 resulted in an identical effect size on *CELSR2*, *PSRC1*, and *SORT1* gene expression. This suggests that the cohort size we have assembled is adequate to accurately perform this particular eQTL analysis in vitro but that there is some aspect of in vitro culture or incomplete HLC maturation that dampens the transcriptional effect of this SNP.

Creation of isogenic genome-edited cell lines in the manner presented here is made possible by previous mechanistic investigation identifying the causal SNP at this locus. In contrast, the in vitro study of other GWAS or eQTL loci will continue to rely on large cohorts of cell lines until their respective causal SNPs are identified. At this time, no genome editing technology is capable of replacing large tracts of genomic DNA, effective in swapping linkage disequilibrium haplotypes in otherwise isogenic backgrounds. Until such a development or identification of causal

SNPs, discovery of and mechanistic investigation into the effect of common genetic variants on cellular phenotypes will be made possible by the use of large iPSC cohorts as presented here.

Although iPSCs and adipocytes lacked *trans*-eQTLs, HLCs also demonstrated a number of *trans*-eQTLs, potentially as downstream effects of decreased lipid accumulation or as direct effects of increased *SORT1* expression. One potential advantage of an in vitro study of common genetic variants is the removal of contaminating cell types from tissue samples, purifying the cell type-specific effect of genotype on gene expression and functional outputs. Although the cohort sizes required for GWASs are orders of magnitude larger than that of this study, we observed that the cohort size of 34 major and 34 minor cell lines was sufficient to accurately model both transcriptional and metabolic effects of this common variant.

Although we assembled this cohort based upon genotype at the rs12740374 SNP, we performed follow-up eQTL analyses on 50 SNPs previously associated with gene expression signatures (Schadt et al., 2008; Teslovich et al., 2010). Four of these loci were marginally significant in correlation with gene expression; DOCK7, GRB14, FRMD5, and HLA-DRB1 correlated with genotype at rs10889353, rs10195252, rs2929282, and rs3177928, respectively (Table S4). These associations were discovered despite the fact that this experiment was underpowered to perform eQTL analysis at loci other than 1p13. Given a larger cohort size, we expect that many additional eQTL analyses and follow-up functional studies will be made possible in HLCs.

Utilizing two metabolomics platforms, we performed intracellular metabolite profiling of iPSCs, adipocytes, and HLCs. Lipid metabolite profiles were dramatically decreased in HLCs carrying the minor 1p13 haplotype but not in cognate iPSCs or adipocytes. The significantly decreased lipid species included phosphatidylethanolamines, triacylglycerols, diacylglycerols, phosphatidylcholines, and sphingomyelins. Further, the results from the Amide metabolite detection platform (a method to quantify predominantly organic acids and similar small molecules) demonstrated a significant reduction of acetoacetic acid in the minor genotype HLCs. The primary function of this ketone body is to provide acetoacetyl CoA and acetyl CoA for synthesis of cholesterol, fatty acids, and complex lipids. These results suggest a globally decreased synthesis of lipid species and a faithful reproduction, in an in vitro cellular model, of the lipid changes seen in the sera of patients with the same hepatic *SORT1* eQTL.

2-aminoadipic acid, a product of lysine degradation, was significantly decreased in minor genotype HLCs. This metabolite has previously been identified by our group as a mediator of cardiometabolic risk (Wang et al., 2013), and it is intriguing that it is decreased in HLCs that carry an eQTL associated with reduced lipid levels. Other metabolomic findings generate more questions than answers. For example, D-gluconic acid, an oxidative product of glucose, was also significantly decreased in minor genotype HLCs. So was kynurenine, a product of tryptophan metabolism and a precursor of niacin, and pseudouridine, a modified nucleoside, but the relevance of these findings is not immediately clear.

These metabolic findings, in addition to the *trans*-eQTLs identified in HLCs, may point toward previously uncharacterized underlying biological mechanisms of MI and other forms of atherosclerotic cardiovascular disease (CVD). One suggestive finding from the *trans*-eQTL gene set is the 3-fold downregulation of *PLTP* (previously associated with plasma lipid concentrations by GWASs) in HLCs carrying the minor 1p13 haplotype compared with major haplotyped HLCs. *PLTP* has been implicated in lipoprotein assembly in hepatocytes (Manchekar et al., 2015; Yazdanyar and Jiang, 2012; Yazdanyar et al., 2013) in a directional manner consistent with our current findings. It is possible that *PLTP* downregulation may affect or contribute to the generally decreased accumulation of lipids in minor haplotype HLCs. Other genes in this dataset have been associated by GWASs with CVD-relevant traits, including *EYS*, *PCDHA9*, *MFSD4B*, *CYB561D1*, *PCDHA10*, *CPN1*, and *SNX5.* Among these genes, *CYB561D1*, *PLTP*, *CPN1*, *CPNE1*, and *SNX5* are expressed in normal human hepatocytes. Future mechanistic studies will delineate the contribution of *PLTP* and these other genes identified by *trans*-eQTL in the relationship between 1p13 haplotype and CVD pathogenesis. Potential mechanistic discoveries of this kind are one of the advantages of an iPSC-driven study of common genetic variants, enabled by the abundance of cellular material made possible by differentiation and the depletion of contaminating cell types prevalent in primary samples.

This study serves as proof of concept that in vitro eQTL analyses are indeed possible using iPSC technology and will facilitate more ambitious population genetics analyses performed in the dish. It is our expectation that improvements in differentiation protocols and reprogramming strategies will make this strategy more accurate and commonplace. Future studies with larger cohorts could be used for functional quantitative trait locus (fQTL) analyses, correlating common genetic variants with cellular functions. Such studies may provide significant incremental functional genomic information regarding tissue-specific mechanisms of disease, thereby extending the utility of traditional QTL and GWAS designs.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
- METHOD DETAILS
  - Peripheral Blood Mononuclear Cell (PBMC) Isolation
  - T Cell Expansion
  - iPSC Reprogramming
  - Immunofluorescence
  - RNA isolation, cDNA synthesis and qPCR
  - Karyotyping
  - EB Scorecard Assay
  - Differentiation into white adipocytes
  - Triglyceride Quantification
  - Differentiation into HLCs
  - FACS
  - Genome Editing
  - APOB ELISA
  - RNA Sequencing

## SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, four tables, and one dataset and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2017.01.010.

## AUTHOR CONTRIBUTIONS

C.R.W. and J.F.O. designed and performed experiments, analyzed data, and prepared the manuscript. M.F. designed and performed experiments. C.E.B. and D.T.P. designed and performed experiments and analyzed data. A.M.T., X.Z., and X.S. analyzed data. J.E.M. performed and analyzed experiments. P.L., Y.W., J.C., F.X., M.H.C.F., E.D., L.C., and J.S. performed experiments. R.C.D. analyzed data and prepared the manuscript. X.J. designed experiments. K.M. supported D.T.P., A.M. supported A.M.T. S.K., L.D., J.Z., R.E.G., R.S.V., and C.J.O. supported and designed the study. C.A.C. designed experiments, prepared the manuscript, and supported the study.

## ACKNOWLEDGMENTS

This study was funded by NIH/NHLBI U01 HL107440 and NIH/NIDDDK R01DK097768-01, start-up funds to C.A.C. from Harvard University, funding from the Harvard Stem Cell Institute, the intramural research program of the National Heart Lung and Blood Institute, NIH, and the NHLBI Framingham Heart Study (contract numbers HHSN268201500001I and N01-HC-25195). The authors would like to thank Dr. Lee Rubin and the Harvard Stem Cell Institute's Therapeutic Screening Center for the use of high-content microscopy devices.

Received: May 18, 2016
Revised: October 10, 2016
Accepted: January 27, 2017
Published: April 6, 2017

## SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Comuzzie et al. (2012); Fox et al. (2012); Inouye et al. (2012); Kettunen et al. (2012); Shin et al. (2014); Willer et al. (2013); Yu et al. (2013); Yuan et al. (2008).

## REFERENCES

Ahfeldt, T., Schinzel, R.T., Lee, Y.K., Hendrickson, D., Kaplan, A., Lum, D.H., Camahort, R., Xia, F., Shay, J., Rhee, E.P., et al. (2012). Programming human pluripotent stem cells into white and brown adipocytes. Nat. Cell Biol. *14*, 209–219.

Aulchenko, Y.S., Ripke, S., Isaacs, A., and van Duijn, C.M. (2007). GenABEL: an R library for genome-wide association analysis. Bioinformatics *23*, 1294–1296.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. A Stat. Soc. *57*, 289–300.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439–452.

Comuzzie, A.G., Cole, S.A., Laston, S.L., Voruganti, V.S., Haack, K., Gibbs, R.A., and Butte, N.F. (2012). Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. PLoS ONE *7*, e51954.

Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T., et al. (2013). A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell *12*, 238–251.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2014. Nucleic Acids Res. *42*, D749–D755.

Fox, C.S., Liu, Y., White, C.C., Feitosa, M., Smith, A.V., Heard-Costa, N., Lohman, K., Johnson, A.D., Foster, M.C., Greenawalt, D.M., et al.; GIANT Consortium; MAGIC Consortium; GLGC Consortium (2012). Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. PLoS Genet. *8*, e1002695.

Freytag, S.O., Paielli, D.L., and Gilbert, J.D. (1994). Ectopic expression of the CCAAT/enhancer-binding protein alpha promotes the adipogenic program in a variety of mouse fibroblastic cells. Genes Dev. *8*, 1654–1663.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Inouye, M., Ripatti, S., Kettunen, J., Lyytikäinen, L.P., Oksala, N., Laurila, P.P., Kangas, A.J., Soininen, P., Savolainen, M.J., Viikari, J., et al. (2012). Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. PLoS Genet. *8*, e1002907.

Kannel, W.B., Feinleib, M., McNamara, P.M., Garrison, R.J., and Castelli, W.P. (1979). An investigation of coronary heart disease in families. The Framingham offspring study. Am. J. Epidemiol. *110*, 281–290.

Kettunen, J., Tukiainen, T., Sarin, A.P., Ortega-Alonso, A., Tikkanen, E., Lyytikäinen, L.P., Kangas, A.J., Soininen, P., Würtz, P., Silander, K., et al. (2012). Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat. Genet. *44*, 269–276.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Lee, Y.K., and Cowan, C.A. (2014). Differentiation of white and brown adipocytes from human pluripotent stem cells. Methods Enzymol. *538*, 35–47.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Manchekar, M., Liu, Y., Sun, Z., Richardson, P.E., and Dashti, N. (2015). Phospholipid transfer protein plays a major role in the initiation of apolipoprotein B-containing lipoprotein assembly in mouse primary hepatocytes. J. Biol. Chem. *290*, 8196–8205.

Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K.V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M., et al. (2010). From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature *466*, 714–719.

Park, I.H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M.W., Cowan, C., Hochedlinger, K., and Daley, G.Q. (2008). Disease-specific induced pluripotent stem cells. Cell *134*, 877–886.

Patsch, C., Challet-Meylan, L., Thoma, E.C., Urich, E., Heckel, T., O'Sullivan, J.F., Grainger, S.J., Kapp, F.G., Sun, L., Christensen, K., et al. (2015). Generation of vascular endothelial and smooth muscle cells from human pluripotent stem cells. Nat. Cell Biol. *17*, 994–1003.

Peters, D.T., Cowan, C.A., and Musunuru, K. (2013). Genome editing in human pluripotent stem cells. In StemBook, L. Girard, ed. (Harvard Stem Cell Institute).

Peters, D.T., Henderson, C.A., Warren, C.R., Friesen, M., Xia, F., Becker, C.E., Musunuru, K., and Cowan, C.A. (2016). Asialoglycoprotein receptor 1 is a specific cell-surface marker for isolating hepatocytes derived from human pluripotent stem cells. Development *143*, 1475–1481.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., and Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. *12*, R22.

Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C., et al. (2008). Mapping the genetic architecture of gene expression in human liver. PLoS Biol. *6*, e107.

Shin, S.Y., Fauman, E.B., Petersen, A.K., Krumsiek, J., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.P., et al.; Multiple Tissue Human Expression Resource (MuTHER) Consortium (2014). An atlas of genetic influences on human blood metabolites. Nat. Genet. *46*, 543–550.

Si-Tayeb, K., Noto, F.K., Nagaoka, M., Li, J., Battle, M.A., Duris, C., North, P.E., Dalton, S., and Duncan, S.A. (2010). Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. Hepatology *51*, 297–305.

Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc. *7*, 500–507.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707–713.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Tsankov, A.M., Akopian, V., Pop, R., Chetty, S., Gifford, C.A., Daheron, L., Tsankova, N.M., and Meissner, A. (2015). A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells. Nat. Biotechnol. *33*, 1182–1192.

Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). Five years of GWAS discovery. Am. J. Hum. Genet. *90*, 7–24.

Wang, T.J., Larson, M.G., Vasan, R.S., Cheng, S., Rhee, E.P., McCabe, E., Lewis, G.D., Fox, C.S., Jacques, P.F., Fernandez, C., et al. (2011). Metabolite profiles and the risk of developing diabetes. Nat. Med. *17*, 448–453.

Wang, T.J., Ngo, D., Psychogios, N., Dejam, A., Larson, M.G., Vasan, R.S., Ghorbani, A., O'Sullivan, J., Cheng, S., Rhee, E.P., et al. (2013). 2-Aminoadipic acid is a biomarker for diabetes risk. J. Clin. Invest. *123*, 4309–4317.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis, First Edition (Springer-Verlag).

Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., and Mora, S.; Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.

Wu, Z., Rosen, E.D., Brun, R., Hauser, S., Adelmant, G., Troy, A.E., McKeon, C., Darlington, G.J., and Spiegelman, B.M. (1999). Cross-regulation of C/EBP alpha and PPAR gamma controls the transcriptional pathway of adipogenesis and insulin sensitivity. Mol. Cell *3*, 151–158.

Yazdanyar, A., and Jiang, X.C. (2012). Liver phospholipid transfer protein (PLTP) expression with a PLTP-null background promotes very low-density lipoprotein production in mice. Hepatology *56*, 576–584.

Yazdanyar, A., Quan, W., Jin, W., and Jiang, X.C. (2013). Liver-specific phospholipid transfer protein deficiency reduces high-density lipoprotein and non-high-density lipoprotein production in mice. Arterioscler. Thromb. Vasc. Biol. *33*, 2058–2064.

Yu, B., Zheng, Y., Alexander, D., Manolio, T.A., Alonso, A., Nettleton, J.A., and Boerwinkle, E. (2013). Genome-wide association study of a heart failure related metabolomic profile among African Americans in the Atherosclerosis Risk in Communities (ARIC) study. Genet. Epidemiol. *37*, 840–845.

Yuan, X., Waterworth, D., Perry, J.R., Lim, N., Song, K., Chambers, J.C., Zhang, W., Vollenweider, P., Stirnadel, H., Johnson, T., et al. (2008). Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. Am. J. Hum. Genet. *83*, 520–528.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| FG-purified Anti-Human CD3, Clone: OKT3 | eBioscience | Cat#160037-81; RRID:AB_468854 |
| OCT4 | Abcam | Cat#ab19857; RRID: AB_445175 |
| NANOG | Abcam | Cat#ab21624; RRID: AB_446437 |
| SSEA3 | Millipore | Cat#MAB4303; RRID: AB_177628 |
| SSEA4 | Millipore | Cat#MAB4304; RRID: AB_177629 |
| Tra-1-60 | Millipore | Cat#MAB4360; RRID: AB_2119183 |
| C/EBP-alpha | Abcam | Cat#ab40761; RRID: AB_726792 |
| Albumin | Bethyl Laboratories | Cat#E80-129; RRID: AB_1286522 |
| ASGR1 | Becton Dickinson Biosciences | Cat#563654 |
| ASGR1-PE | Becton Dickinson Biosciences | Cat#563655 |
| Mouse IgGκ Isotype Control | Becton Dickinson Biosciences | Cat#555749; RRID: AB_396091 |
| Biological Samples | | |
| Human Blood Samples | Framingham Heart Study Offpsring Cohort | N/A |
| Chemicals, Peptides, and Recombinant Proteins | | |
| Human Recombinant Interleukin 2 | Becton Dickinson | Cat#354043 |
| Collagenase IV | Stem Cell Technologies | Cat#07909 |
| CHIR99021 | Cayman Chemical | Cat#13122-10mg |
| Activin A | Cell Guidance Systems | Cat#GFH6-1000 |
| Bone Morphogenetic Protein 4 | R&D Systems | Cat#314-BP |
| Hepatocyte Growth Factor | Peprotech | Cat# 100-39 |
| Dimethyl Sulfoxide | Sigma-Aldrich | Cat#D2650 |
| Oncostatin M | Peprotech | Cat#300-10 |
| Dexamethasone | Sigma-Aldrich | Cat#D1756 |
| Rosiglitazone | Santa Cruz | Cat#sc-202795 |
| Insulin | Sigma-Aldrich | Cat#I9278 |
| Basic Fibroblast Growth Factor | Aldevron | Cat#7001 |
| Critical Commercial Assays | | |
| Cytotune-iPS 2.0 Sendai Reprogramming Kit | Thermo Fisher | Cat# A16517 |
| RNeasy Mini Kit | QIAGEN | Cat# 74104 |
| High-Capacity cDNA Reverse Transcription Kit | Thermo Fisher | Cat#4368814 |
| 384-well TaqMan hPSC Scorecard kits | Thermo Fisher | Cat#A15872 |
| *DC* protein assay | Bio-Rad Laboratories | Cat# 5000116 |
| Infinity Triglycerides Reagent | VWR | Cat# 46100-346 |
| Human APOB ELISAPro kit | Mabtech | Cat# 3715-1HP-2 |
| GLOBINclear Kit, human, for globin mRNA depletion | Thermo Fisher | AM1980 |
| TruSeq RNA Library Prep Kit v2 | Illumina | Cat#RS-122-2001 |
| Deposited Data | | |
| RNA Sequencing data | NCBI Gene Expression Omnibus | GEO: GSE90749 |
| Experimental Models: Cell Lines | | |
| 69 iPSC lines | This Study | N/A |
| rs12740374 knockout clones | This Study | N/A |
| Recombinant DNA | | |
| FU-TetO-Gateway | Addgene | Cat#43914 |

(*Continued on next page*)

**Continued**

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Lenti-Doxycycline-PPARG2 | (Ahfeldt et al., 2012; Lee and Cowan, 2014) | N/A |
| FUdeltaGW-rtTA | Addgene | Cat#19780 |
| pRSV-Rev | AddGene | Cat#12253 |
| pMDLg/pRRE | AddGene | Cat#12251 |
| pCas9_GFP | Addgene | Cat#44719 |
| gRNA cloning Vector | Addgene | Cat#41824 |
| Sequence-Based Reagents | | |
| *RPLPO* Forward Primer (5′ to 3′): GCAGCATCTACAACCCTGAAG | This Study | N/A |
| *RPLPO* Reverse Primer (5′ to 3′): GCAGACAGACACTGGCAACA | This Study | N/A |
| *DNMT3B* Forward Primer (5′ to 3′): ATAAGTCGAAGGTGCGTCGT | This Study | N/A |
| *DNMT3B* Reverse Primer (5′ to 3′): GGCAACATCTGAAGCCATTT | This Study | N/A |
| *hTERT* Forward Primer (5′ to 3′): TGTGCACCAACATCTACAAG | This Study | N/A |
| *hTERT* Reverse Primer (5′ to 3′): GCGTTCTTGGCTTTCAGGAT | This Study | N/A |
| *NANOG* Forward Primer (5′ to 3′): CAGTCTGGACACTGGCTGAA | This Study | N/A |
| *NANOG* Reverse Primer (5′ to 3′): CTCGCTGATTAGGCTCCAAC | This Study | N/A |
| *hTERT* Forward Primer (5′ to 3′): TGTGCACCAACATCTACAAG | This Study | N/A |
| *hTERT* Reverse Primer (5′ to 3′): GCGTTCTTGGCTTTCAGGAT | This Study | N/A |
| *OCT4* Forward Primer (5′ to 3′): TGTACTCCTCGGTCCCTTTC | This Study | N/A |
| *OCT4* Reverse Primer (5′ to 3′): TCCAGGTTTTCTTTCCCTAGC | This Study | N/A |
| *REX1* Forward Primer (5′ to 3′): TGGACACGTCTGTGCTCTTC | This Study | N/A |
| *REX1* Reverse Primer (5′ to 3′): GTCTTGGCGTCTTCTCGAAC | This Study | N/A |
| *SOX2* Forward Primer (5′ to 3′): GCTAGTCTCCAAGCGACGAA | This Study | N/A |
| *SOX2* Reverse Primer (5′ to 3′): GCAAGAAGCCTCTCCTTGAA | This Study | N/A |
| *NT5E* Gene Expression Assay | Thermo Fisher | Cat#Hs00159686_m1 |
| *CD44* Gene Expression Assay | Thermo Fisher | Cat#Hs01075861_m1 |
| *ENG* Gene Expression Assay | Thermo Fisher | Cat#Hs00923996_m1 |
| *CD4* Gene Expression Assay | Thermo Fisher | Cat#Hs01058407_m1 |
| *ALB* Gene Expression Assay | Thermo Fisher | Cat#Hs00910225_m1 |
| *ADIPOQ* Gene Expression Assay | Thermo Fisher | Cat#Hs00605917_m1 |
| *FABP4* Gene Expression Assay | Thermo Fisher | Cat#Hs01086177_m1 |
| *APOB* Gene Expression Assay | Thermo Fisher | Cat#Hs01071209_m1 |
| *SARS* Gene Expression Assay | Thermo Fisher | Cat#Hs00197856_m1 |
| *CELSR2* Gene Expression Assay | Thermo Fisher | Cat#Hs00154903_m1 |
| *PSRC1* Gene Expression Assay | Thermo Fisher | Cat#Hs00934027_g1 |
| *MYBPHL* Gene Expression Assay | Thermo Fisher | Cat#Hs00975401_m1 |
| *SORT1* Gene Expression Assay | Thermo Fisher | Cat#Hs00361760_m1 |
| *PSMA5* Gene Expression Assay | Thermo Fisher | Cat#Hs00936004_m1 |
| *RPLP0* Gene Expression Assay | Thermo Fisher | Cat#326314E |
| rs12740374 Genotyping F (5′ to 3′): AGTTGCTGACCCAAAAGTGC | This Study | N/A |
| rs12740374 Genotyping R (5′ to 3′): GAGGCCACAGCAGGTTAGAC | This Study | N/A |
| Software and Algorithms | | |
| GraphPad Prism | Graphpad Software | N/A |
| Bowtie2 | (Kim et al., 2013) | http://bowtie-bio.sourceforge.net/bowtie2/index.shtml |
| Cufflinks | (Roberts et al., 2011) | http://cole-trapnell-lab.github.io/cufflinks/ |
| Subread package for R | (Liao et al., 2014) | http://subread.sourceforge.net/ |

(*Continued on next page*)

| Continued | | |
|---|---|---|
| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
| PEER package for R | (Stegle et al., 2012) | http://www.sanger.ac.uk/science/tools/peer |
| Multiquant Software | Sciex | https://sciex.com |
| GenABEL package for R | (Aulchenko et al., 2007) | https://cran.r-project.org/web/packages/GenABEL/index.html |
| Other | | |
| Vacutainer CPT Cell Preparation Tubes with Sodium Citrate, 4 mL draw capacity | Becton Dickinson | Cat#362760 |
| X-VIVO 10 with Gentamicin and Phenol Red, 1 L | Lonza | Cat#04-380Q |
| 5% Type AB Heat Inactivated Human Serum | Valley Biomedical | Cat#HP1022HI |
| mTeSR 1 | Stem Cell Technologies | Cat#05850 |
| Gentle Cell Dissociation Reagent | Stem Cell Technologies | Cat#07174 |
| Knockout Serum Replacement | Thermo Fisher | Cat#10828010 |
| LipidTox Red and LipidTox Deep Red dyes | Thermo Fisher | Cat# H34476 and H34477 |
| Taqman Gene Expression Master Mix | Thermo Fisher | Cat# 4369016 |
| FAST SYBR Green Master Mix | Thermo Fisher | Cat# 4385616 |
| Accutase | Stem Cell Technologies | Cat# 07920 |
| Y-27632 | Stem Cell Technologies | Cat# 72302 |
| Methanol (Optima LC/MS) | Fisher Chemical | Cat#A456-1 |
| 2-Propanol (Optima LC/MS) | Fisher Chemical | Cat#A461-1 |
| Water (HPLC) | Fisher Chemical | W5-4 |
| Chloroform (HPLC) | Fisher Chemical | C607-1 |
| L-Phenylalanine-$^{13}C_9$,$^{15}N$,α,β,β,2,3,4,5,6-$d_8$ | Sigma Aldrich | *Cat#749885* |
| L-Valine-$d_8$ | Cambridge Isotope Laboratories | Cat# DLM-7784-PK |
| Acetonitrile, HPLC | J.T. Baker | Cat# 9012-03 |
| XBridge BEH Amide Column, 130Å, 3.5 μm, 4.6 mm X 30 mm, 1/pkg | Waters | Cat#186004866 |
| Ammonium Acetate | Sigma Aldrich | Cat#17836-50G |
| Ammonium hydroxide solution | Sigma Aldrich | Cat#*17837* |

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by, the Lead Contact, Chad Cowan (chad_cowan@harvard.edu).

## EXPERIMENTAL MODELS AND SUBJECT DETAILS

iPSC lines were derived from participants of the FHS Offspring Cohort. For details concerning the sex and karyotype of each cell line, refer to Table S1. Informed consent was obtained from all participants under Institutional Review Board Approval at both Boston University and Harvard University. All iPSC lines will be made available through WiCell in line with standard policies for academic and corporate users.

## METHOD DETAILS

### Peripheral Blood Mononuclear Cell (PBMC) Isolation

Fasting blood specimens were collected on the morning of PBMC isolation. Samples were placed in Vacutainer® CPT Cell Preparation Tubes with Sodium Citrate, 4 mL Draw Capacity (BD) with PBS added to a volume of 3.5 mL, and centrifugation and cell washing steps were performed according to the manufacturer's instructions. Isolated PBMCs were resuspended in 1 mL freezing media (90% FBS, 10% DMSO) and pipetted into cryovials. Samples were frozen at −80°C for 24 hr using a Mr. Frosty Freezing Container (Thermo Fisher) then stored in liquid nitrogen. All samples were isolated, processed and cryopreserved within four hours of collection.

### T Cell Expansion

PBMCs (approximately $1x10^7$ cells per sample) were thawed into one well of a 12-well plate. To promote isolation of T cells, this plate was pre-coated with 10 μg/mL FG Purified Anti-Human CD3 in PBS for 1 hr at 37°C, blocked using 2% BSA in PBS for 30 min at 37°C and rinsed twice with PBS. Before reprogramming, cells were cultured for five days in T cell expansion media (X-VIVO 10 with Gentamicin and Phenol Red (Lonza), 5% Type AB Heat Inactivated Human Serum (Valley Biomedical), 1% Pen Strep, 0.4 μL/mL FG Purified Anti-Human CD3, 0.4 μL/mL FG Purified Anti-Human CD28 (eBioscience), and 100 U/mL human recombinant Interleukin-2 (BD)). All cells were maintained in 5% $CO_2$ incubators at 37°C.

### iPSC Reprogramming

On day zero, $1.5x10^5$ T cells were seeded in 700 μL T cell expansion media and transduced by adding vectors from a CytoTune-iPS Sendai Reprogramming Kit (MOI Oct3/4:Sox2:Klf4:cMyc = 10:10:10:10) or CytoTune-iPS 2.0 Sendai Reprogramming Kit (MOI KOS: hc-Myc: hKlf4 = 6:6:3.6) (Thermo Fisher). After twenty-four hours, cells were re-plated in 1 mL of mouse embryonic fibroblast (MEF media (DMEM, 10% FBS, 1% Pen Strep, 1% L-Glutamine). After forty-eight hours, cells were transferred to a 10-cm dish pre-seeded with $2x10^6$ irradiated MEFs on 0.1% gelatin. Seventy-two hours later, cells were fed with MEF media. For the first four media changes, spent media were centrifuged to recover non-adherent cells. From day seven for at least three weeks, cells were switched to hESC media (DMEM/F12, 20% KnockOut Serum Replacement (Thermo Fisher), 1% L-Glutamine, 1% Pen Strep, 1% MEM-NEAA, 1 μL/mL 55 mM 2-Mercaptoethanol, 10 ng/mL bFGF). Cells were fed every other day until day 14, then every day thereafter. After approximately three weeks of feeding with hESC medium, colonies were picked into separate wells pre-coated with $1.8x10^4/cm^2$ irradiated MEFs on 0.1% gelatin. iPSCs on MEFs were cultured in hESC media and passaged as aggregates using Collagenase IV for 10 min at 37°C with manual scraping. After approximately four weeks, iPSCs were transferred to plates coated with hESC-qualified Matrigel (Corning) and cultured in mTeSR1 (StemCell Technologies, Inc.). Feeder-free iPSCs were passaged as aggregates with Gentle Cell Dissociation Reagent (StemCell Technologies, Inc.) for five minutes at room temperature with manual scraping.

### Immunofluorescence

Cells were fixed in ice-cold methanol for fifteen minutes at minus twenty degrees Celsius (°C). After washing with phosphate buffered saline (PBS), cells were incubated in blocking buffer (PBS supplemented with 5% bovine serum albumin, 0.1% Triton X-100) for one hour at room temp (RT). All primary antibody incubations occur overnight at four °C. The following day, cells were washed with PBS three times, then incubated with secondary antibody in blocking buffer for one hour at RT in the dark. After this incubation, buffer was decanted and replaced with PBS supplemented with Hoechst and neutral lipid dye at RT for thirty minutes in the dark. Cells were then washed twice more with PBS. Primary and secondary antibodies listed in the Key Resources Table. The LipidTox Red and LipidTox Deep Red dyes were used for labeling neutral lipids (Thermo Scientific).

### RNA isolation, cDNA synthesis and qPCR

RNA was extracted from adherent cells using TRIzol® reagent or from FACS-sorted cells using TRIzol® LS reagent according to the manufacturer's instructions (Thermo Scientific). cDNA synthesis was performed using qScript cDNA Supermix according to the manufacturer's instructions (Quanta Biosciences). SeV detection qPCR, *NT5E, CD44, ENG, CD4, FABP4, ADIPOQ, APOB, ALB, SARS, CELSR2, PSRC1, MYBPHL, SORT1, PSMA5,* and *RPLP0* qPCR were performed using TaqMan Expression Assays and TaqMan Gene Expression Mastermix on a Viia7 Real-Time PCR System (Thermo Scientific). TaqMan Gene Expression Assays are listed in the Key Resources Table. *DNMT3B, TERT, NANOG, OCT4, REX1,* and *SOX2* qPCR was performed using primer sets listed in the Key Resources Table in conjunction with Fast SYBR® Green master mix (Thermo Scientific). Expression data is presented after calculating the relative expression compared with the housekeeping gene *RPLP0*, using the equation Relative Quantification (RQ) = 100/(2^(Target Gene Ct – *RPLP0* Ct).

### Karyotyping

Karyotype G-banding was performed by the Tufts Medical Center Cytogenetics Laboratory (Tufts University) at least one passage after the Sendai virus was no longer detectable in the iPSC population by qPCR.

### EB Scorecard Assay

To form embryoid bodies (EBs), iPSCs were passaged as aggregates using Accutase (diluted 1:3 in PBS) for three minutes at 37°C, transferred to ultra-low attachment plates and cultured for seven days in EB formation media (DMEM, 10% KnockOut Serum Replacement, 1% Pen Strep). 0.44 μL/mL 10 mM Y-27632 (StemCell Technologies, Inc.) was included for the first 48 hr of culture. Media was changed every other day. EBs were then re-plated on 0.1% gelatin-coated plates in DMEM with 10% FBS for an additional seven days, changing media every other day, then harvested using a cell scraper. RNA was isolated from EBs using RNeasy Mini Kit (QIAGEN) and cDNA was prepared using High Capacity cDNA Reverse Transcription Kit (Thermo Fisher). Gene expression was analyzed using 384-well TaqMan® hPSC Scorecard kits (Thermo Fisher).

### Differentiation into white adipocytes

We differentiated white adipocytes following our published protocol (Ahfeldt et al., 2012; Lee and Cowan, 2014). Briefly, MPCs were generated by EB formation in low-adherence tissue culture dishes followed by EB plating onto plates coated with 1% gelatin. After

propagation of the MPC subpopulation by selective adhesion to gelatin, MPCs were transduced with Tet-On *PPARG2* and rtTA lentiviruses. After growing to confluency, doxycycline was added in adipogenic medium for sixteen days. Cells were harvested after culture for an additional six days in doxycycline-free adipogenic medium.

### Triglyceride Quantification

White adipocytes were washed with PBS and then harvested in 5% nonidet-P-40 in water. Prior to triglyceride quantification, cells were homogenized by bath sonication. Aliquots were diluted one to ten and protein concentration was quantified by *DC* protein assay (Bio-Rad Laboratories). Samples were then heated and vortexed to solubilize triglycerides, and triglycerides were quantified by Infinity Triglycerides reagent according to the manufacturer's instructions (VWR). Triglyceride concentrations were interpolated to a triglyceride standard curve (Abcam).

### Differentiation into HLCs

HLCs were generated using our published directed differentiation protocol (Peters et al., 2016). Briefly, Activin A and CHIR99021 treatment resulted in DE, Bone Morphogenetic Protein 4 and Fibroblast Growth Factor-2 treatment produced HE, Hepatocyte Growth Factor (HGF) treatment yielded IMH, and these cells were matured in Hepatocyte basal medium supplemented with HCM Singlequot Kits (Lonza), HGF, Oncostatin M, and Dexamethasone.

### FACS

Mature HLCs were isolated following our published protocol (Peters et al., 2016). We used FACS to isolate mature HLCs by tagging surface ASGR1 with an R-Phycoerythrin (PE)-labeled antibody (Becton Dickinson). PE-tagged isotype controls were used for each differentiation replicate for each cell line, to correct for background fluorescence and non-specific binding of the antibody. FACS was performed using a FACS Aria (Becton Dickinson Biosciences) fitted with an eighty-five μm nozzle.

### Genome Editing

CRISPR/Cas9 genome editing of iPSCs was carried out according to our published protocol (Peters et al., 2013). CRISPR single guide RNAs (sgRNAs) were cloned into a publically available sgRNA cloning vector and are shown in Figure S3A. Clonal cell lines were derived by limiting dilution and were screened for mutagenesis by PCR. Oligonucleotide primers used for mutagenesis screening are listed in the Key Resources Table. Double knockout of rs12740374 in these clonal cell lines was verified by Sanger sequencing (Genewiz).

### APOB ELISA

HLC conditioned medium was collected after twenty-four hours in culture and stored at minus eighty °C until the ELISA was performed. The Human APOB ELISAPro kit (Mabtech) were used according to the manufacturer's instructions.

### RNA Sequencing

Total RNA was isolated and extracted using the Qiacube automated sample prep system (QIAGEN); most globin RNA was removed (GLOBINclear Kits, Life Technologies, and Grand Island, NY, USA). 100ng of total RNA was used as starting materials. Illumina TrueSeq kits (Illumina, Inc., La Jolla, CA, USA) were used for RNA purification, chemical fragmentation, single-stranded cDNA conversion, DNA library preparation and oligonucleotide barcoding, which generates 250-450bp library fragments for sequencing with a non-polyA selection library construction (NuGen kit). Sequencing was carried out with Illumina HiSeq 2000 (Illumina, Inc.; 75-bp pair-ended reads, 6 library/samples per lane) for 204 samples (68 samples for iPSC, adipocytes, HLCs, respectively), yielding 65.0 million unique mapping reads (i.e., ~32.5M paired-end reads (average) per sample, with 54.8 million reads mapped in proper pairs.

### Metabolomics

Media was removed from dishes and the plates immersed in high performance liquid chromatography (HPLC) - grade water, the water decanted and aspirated, and then the culture plate carefully placed on the surface of liquid nitrogen in a liquid nitrogen-resistant basin for 15 s to quench cellular metabolic activity. Next, 2mls of extraction medium (Amide Method: 75% 9:1 Methanol (Sigma-Aldrich): Chloroform (Sigma Aldrich), 25% $H_2O$; Lipid Method: 2-Propanol (Fisher Scientific)) were poured into each well. After 15 s 1.5ml was aspirated into respective labeled sample tubes, and 0.5ml was aspirated from each well into collective pooled extract tubes. Extracts were then centrifuged (14,000 rpm, 4°C for 20 min), and 1.25ml of the supernatant was aliquoted into conical Eppendorf tubes and labeled isotope standards (L-Phenylalanine-$d_8$ [Sigma-Adrich] and L-Valine-$d_8$ [Cambridge Isotope Laboratories]) were added to the supernatants. Samples were then dried down on a speedvac concentrator (Thermo Scientific) and re-suspended in 100 μL of acetonitrile (J.T. Baker): water (J.T. Baker) (50: 50 v/v) before injection. Sample injection volume was 5 or 10 μl, depending on liquid chromatography-tandem mass spectrometry (LC-MS) acquisition method, described below.

LC-MS data were acquired using two methods on two LC-MS machines. The first method, the Lipid method, was acquired using a 4000 QTrap triple quadrupole mass spectrometer (Applied Biosystems/Sciex) that was coupled to a multiplexed LC system comprised of a 1200 Series pump (Agilent Technologies) and an HTS PAL autosampler (Leap Technologies) equipped with two injection ports and a column selection valve, as previously reported (Wang et al., 2011). Cellular lipid extracts were analyzed using a 150 mm x 3.2 mm Phosphere C4 column (Grace) and mobile phases (mobile phase A: 10mM Ammonium Acetate [Sigma

Aldrich]: Methanol [JT Baker]: Acetic Acid [Sigma Aldrich], 95:5:0.1 v/v/v, mobile phase B: 0.1% Acetic acid [Sigma Aldrich] in Methanol [JT Baker]). A 10 μL volume of extract was injected directly onto the column under initial conditions (80:20 Mobile Phase A: Mobile Phase B, with a 350 μL/min flow rate). The solvent composition was held constant for 2 min, followed by a linear gradient to 20:80 A:B over one minute, and a further linear gradient to 0:100 A:B over 12 min, where it was held for 10 min before returning to initial conditions for a 10 min re-equilibration. The LC system was connected to an API-4000 QTrap triple quadrupole mass spectrometer (Applied Biosystems / AB Sciex) run in positive ion mode. MS ionization was achieved using an electrospray ionization (ESI) source. Ions were measured in full scan mode (Mass range: 400-1100 Daltons, Dwell time: 1.25 s / scan for a total of 1099 scans / sample). Metabolite peaks were integrated using Multiquant Software (AB Sciex).

The second acquisition method, the Amide method, was optimized for nucleotide and nucleoside phosphates, high-energy intermediates, organic acids, TCA cycle intermediates, and glycolytic intermediates. This method used HILIC chromatography on a 2.1 × 100mm 3.5 μm Xbridge Amide column (Waters) in negative ion mode. Mobile phase A was 95:5 (v/v) water (Fisher Scientific: acetonitrile (Fisher Scientific) with 20 mM ammonium acetate (Sigma-Aldrich) and 20mM ammonium hydroxide (Sigma-Aldrich) (pH 9.5). Mobile phase B was acetonitrile (Fischer Scientific). The chromatography system consisted of a 1260 Infinity autosampler (Agilent) connected to a 1290 Infinity HLPC binary pump system (Agilent). Injection volume was 5 ul. The initial conditions were 0.25 mL/min of 85% mobile phase B followed by a linear gradient to 35% mobile phase B over 6 min. This was followed by a linear gradient to 2% mobile phase B over 0.5 min held for an additional 0.5 min, then a 0.5 min gradient return to 85% mobile phase B. Column equilibration was continued for 4.5 min at 0.5 mL/min for a total cycle time of 12 min. The column compartment was maintained at 30°C. The HPLC pump was connected to a 6490 QQQ (Agilent) triple quadrupole mass spectrometer equipped with an electrospray ionization source, on which 157 metabolites were optimized for negative mode detection. 59 of the metabolites were reproducibly detected in cell extract across all cell types and were selected for dynamic multiple reaction monitoring (MRM) that had a minimum dwell time of 30 ms for each metabolite. Final mass spectrometry settings for the QQQ 6490 were sheath gas temperature 400°C, sheath gas flow 12 L/min, drying gas temperature 290°C, drying gas flow 15 L/min, capillary voltage 4000V, nozzle pressure 30 psi, nozzle voltage 500V, and delta EMV 200V. Metabolite quantification was determined by integrating peak areas using MassHunter QQQ Quant (Agilent).

## QUANTIFICATION AND STATISTICAL ANALYSES

### Analysis of qPCR and ELISA data

Prism 5 (Graphpad) was used to create charts and perform statistical analyses in Figures 1, 2, and 3. Correlation analyses were performed using the Pearson Correlation algorithm. One-sided t tests were performed to analyze the in vitro eQTL as measured by qPCR.

### Analysis of RNA-Sequencing Data

Data quality of raw sequencing data (.fastq) for each sample was assessed using FASTQC, and 75bp paired-reads were mapped to the NCBI v37 *Homo sapiens* reference genome (hg19) using Bowtie2 within Tophat2 (Kim et al., 2013). After alignment, using an annotation file (Ensembl v74) (Flicek et al., 2014), FPKM values were derived using Cufflinks (Roberts et al., 2011) to be used as expression measurements for each feature. One adipocyte sample was sequenced twice. The technical replicate with lower number of sequenced reads and unique mapping rate was removed. After assessment of quality based on overall mapping rate and outlier detection by principle component analysis (PCA), all 204 samples remained for use in the downstream analysis of this study. Association of normalized, log transformed gene level FPKM values with genotype was computed using a linear regression within each cell type, respectively.

Gene ontology analyses were performed using RNA-sequencing data. Using log transformed gene-level FPKM expression values, we first conducted linear regression statistical analyses to identify differentially expressed genes (DEG) between IPSCs, adipocytes and HLCs. Then we performed the gene ontology (GO) enrichment analysis using DAVID (Huang et al., 2009) for these DEGs.

For linear mixed model analyses of RNA-sequencing, reads were mapped to the human genome (hg19) using TopHat (Trapnell et al., 2009). Gene level counts were generated using the featureCounts program from the Subread package (Liao et al., 2014). Data were normalized using the DESeq2 package with application of a variance stabilizing transform. Hidden factors contributing to variance were computed using the PEER package (Stegle et al., 2012). Association of normalized, transformed gene level counts with genotype were computed using linear regression with adjustment for participant identifier (given that there were technical replicates) and the first 10 factors generated by PEER. P values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg false discovery control procedure (Benjamini and Hochberg, 1995).

### Analysis of Metabolomics Data

For both methods, all metabolite peaks were manually reviewed for peak quality in a blinded manner. In addition, pooled cellular extracts were run every 10 injections, enabling the monitoring and correction for temporal drift in mass spectrometry performance. All samples were normalized to the nearest pooled sample in a metabolite-by-metabolite manner. Internal standard peak areas were monitored for quality control and individual samples with peak areas differing from the group mean by more than two standard deviations were re-analyzed.

Normalized metabolite counts were rank normalized using the GenABEL package (Aulchenko et al., 2007) and tested for association with genotype in a linear regression model with adjustment for analysis batch, participant identifier, replicate number, and date of harvest. P values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg false discovery control procedure (Benjamini and Hochberg, 1995).

Data were plotted using the ggplot2 package in R (Wickham, 2009).

All R analyses can be viewed in Data S1.

## DATA AVAILABILITY

The accession number for the RNA-sequencing datasets reported in this paper is GEO: GSE90749.