

## RESEARCH ARTICLE

# Predictive network modeling in human induced pluripotent stem cells identifies key driver genes for insulin responsiveness

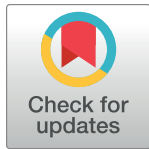
Ivan Carcamo-Orive<sup>1</sup>✉\*, Marc Y. R. Henrion<sup>2,3,4</sup>✉, Kuixi Zhu<sup>2,5,6</sup>✉, Noam D. Beckmann<sup>2</sup>, Paige Cundiff<sup>7</sup>, Sara Moein<sup>2,5,6</sup>, Zenan Zhang<sup>2</sup>, Melissa Alamprese<sup>5,6</sup>✉, Sunita L. D'Souza<sup>8</sup>, Martin Wabitsch<sup>9</sup>, Eric E. Schadt<sup>10</sup>, Thomas Quertermous<sup>1</sup>✉, Joshua W. Knowles<sup>1</sup>‡\*, Rui Chang<sup>2,5,6,11</sup>‡\*

**1** Stanford University School of Medicine, Division of Cardiovascular Medicine, Cardiovascular Institute, and Diabetes Research Center, Stanford, California, United States of America, **2** Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, **3** Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, United Kingdom, **4** Malawi—Liverpool—Wellcome Trust Clinical Research Programme, Blantyre, Malawi, **5** Department of Neurology, University of Arizona, Tucson, Arizona, United States of America, **6** The Center for Innovations in Brain Sciences, University of Arizona, Tucson, Arizona, United States of America, **7** Vertex Pharmaceuticals, Boston, Massachusetts, United States of America, **8** Department of Cellular, Developmental and Regenerative Biology, Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, **9** Department of Pediatrics and Adolescent Medicine, Division of Pediatric Endocrinology, Ulm University, Ulm, Germany, **10** Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, **11** INTelico Therapeutics LLC, Tucson, Arizona, United States of America

✉ These authors contributed equally to this work.

‡ senior author.

\* [ivantxo@stanford.edu](mailto:ivantxo@stanford.edu) (ICO); [knowlej@stanford.edu](mailto:knowlej@stanford.edu) (JWK); [ruichang@email.arizona.edu](mailto:ruichang@email.arizona.edu) (RC)



## OPEN ACCESS

**Citation:** Carcamo-Orive I, Henrion MYR, Zhu K, Beckmann ND, Cundiff P, Moein S, et al. (2020) Predictive network modeling in human induced pluripotent stem cells identifies key driver genes for insulin responsiveness. *PLoS Comput Biol* 16(12): e1008491. <https://doi.org/10.1371/journal.pcbi.1008491>

**Editor:** Sriram Chandrasekaran, University of Michigan, UNITED STATES

**Received:** December 10, 2019

**Accepted:** November 3, 2020

**Published:** December 23, 2020

**Copyright:** © 2020 Carcamo-Orive et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** RNA-seq data is deposited at GEO: GSE79636 and dbGAP: phs001139.

**Funding:** This work was funded by the NIH through P30DK116074 (TQ and JWK), U01HL107388-01 (TQ), 1RF1AG057457-01 (RC), DK116750 (JWK), DK120565 (JWK), DK106236 (JWK), DK107437 (JWK), U41HG009649 (JWK), and ADA 1-19-JDF-108 (JWK). The funders had no role in study design, data collection and analysis, decision to

## Abstract

Insulin resistance (IR) precedes the development of type 2 diabetes (T2D) and increases cardiovascular disease risk. Although genome wide association studies (GWAS) have uncovered new loci associated with T2D, their contribution to explain the mechanisms leading to decreased insulin sensitivity has been very limited. Thus, new approaches are necessary to explore the genetic architecture of insulin resistance. To that end, we generated an iPSC library across the spectrum of insulin sensitivity in humans. RNA-seq based analysis of 310 induced pluripotent stem cell (iPSC) clones derived from 100 individuals allowed us to identify differentially expressed genes between insulin resistant and sensitive iPSC lines. Analysis of the co-expression architecture uncovered several insulin sensitivity-relevant gene sub-networks, and predictive network modeling identified a set of key driver genes that regulate these co-expression modules. Functional validation in human adipocytes and skeletal muscle cells (SKMCs) confirmed the relevance of the key driver candidate genes for insulin responsiveness.

publish, or preparation of the manuscript. JWK, I C-O and TQ received salary support through the stated NIH grants.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: RC is the founder of INTelico Therapeutics LLC.

## Author summary

Insulin resistance is characterized by a defective response (“resistance”) to normal insulin concentrations to uptake the glucose present in the blood, and is the underlying condition that leads to type 2 diabetes (T2D) and increases the risk of cardiovascular disease. It is estimated that 25–33% of the US population are insulin resistant enough to be at risk of serious clinical consequences. For more than a decade, large population studies have tried to discover the genes that participate in the development of insulin resistance, but without much success. It is now increasingly clear that the complex genetic nature of insulin resistance requires novel approaches centered in patient specific cellular models. To fill this gap, we have generated an induced pluripotent stem cell (iPSC) library from individuals with accurate measurements of insulin sensitivity, and performed gene expression and key driver analyses. Our work demonstrates that iPSCs can be used as a revolutionary technology to model insulin resistance and to discover key genetic drivers. Moreover, they can develop our basic knowledge of the disease, and are ultimately expected to increase the therapeutic targets to treat insulin resistance and type 2 diabetes.

## Introduction

Insulin resistance is necessary for the development of the metabolic syndrome and type 2 diabetes (T2D), and is a major risk factor for cardiovascular disease [1], which together represent a modern pandemic. While genome-wide association studies (GWAS) have identified a large number of genomic loci associated with T2D-related traits, most of these signals are associated with pancreatic  $\beta$ -cell function and insulin secretion rather than with insulin resistance [2]. While a few insulin resistance genes have been identified [3–6], the underlying genetic architecture of insulin resistance remains unknown.

To fill this gap, we sought to take advantage of a large library of induced pluripotent stem cells (iPSCs) derived from individuals across the spectrum of insulin sensitivity who have also undergone GWAS genotyping [7,8]. We have fully characterized these iPSC lines and demonstrated determinants of iPSC transcriptional variability. For instance, we found that the highest across individual contribution to variability in our cohort was enriched for metabolic functions [9].

These results prompted us to more specifically analyze the gene expression patterns and networks associated with the insulin sensitivity status of the iPSC donors. For complex conditions like insulin resistance with polygenic susceptibility, systems biology and network modeling, integrating multiscale-omics data like genetic and transcriptomic data, provide a useful context in which to interpret associations between genes and functional variation or disease states [9–13]. Therefore, the reconstruction of molecular networks can lead to a more systematic and data driven characterization of pathways underlying disease, and consequently, a more comprehensive approach to identifying and prioritizing therapeutic targets [12,13]. Recent advances in co-expression and causal/predictive network modeling [9,11,12,14] allow us to take such an approach. The work described here links complex disease phenotypes from highly characterized subjects to concomitant molecular networks that can then be used to uncover coherent, functional molecular sub-networks and their key driver genes that ultimately determine the clinical phenotypes.

In summary, we performed differential expression analyses between insulin resistant (IR) and insulin sensitive (IS) iPSCs, built co-expression networks to systematically organize the data into coherent modules enriched for insulin sensitivity associated functions and

implemented key driver analyses to identify genes that control and regulate critical aspects of the IR and IS networks. Finally, we empirically validated the constructed networks in iPSCs and the prioritized key drivers through insulin responsiveness associated functional assays in human adipocytes and skeletal muscle cells (SKMCs)([Fig 1](#)).

## Results

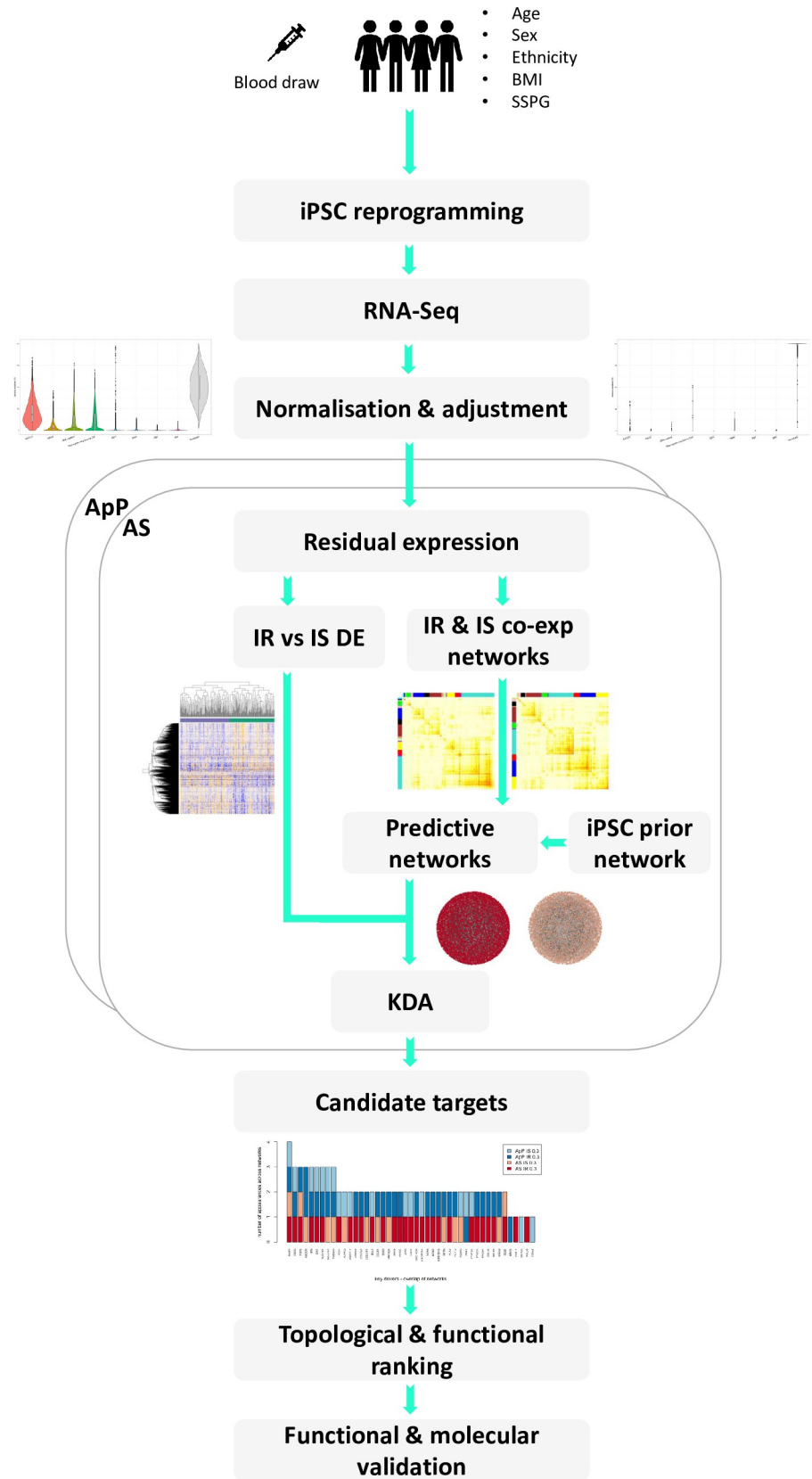
### Insulin sensitivity measurement and iPSC generation

Individuals in the study have accompanying genome-wide genotyping and gold standard measurement of insulin sensitivity (i.e. steady state plasma glucose-SSPG-derived from an insulin suppression test). Other biometric parameters include age, body mass index, sex and race/ethnicity ([S1 Table](#)). We generated three to seven iPSC lines from each individual, with no apparent differences in the reprogramming efficiency between IR and IS cells ([S1A Fig](#)). The complete pipeline for iPSC generation and quality control has been previously described [9]. Briefly, we generated RNA-seq data for 317 iPSC lines from 101 individuals and, after quality control [9], we analyzed RNA-seq data from 310 samples from 100 individuals, of which 48 were IS (149 samples) and 52 were IR (161 samples). The SSPG cut-off to discriminate between IR or IS state was set at 140 mg/dl based on previous publications [15–17] ([S1B Fig](#)). The average SSPG values were 84 mg/dl for the IS group and 210 mg/dl for the IR group. Finally, samples in both groups were age and body mass index (BMI) matched to avoid possible biases (mean age 57.7 years old and 59.5 in the IS vs. IR group, respectively, and mean BMI 28.5 in the IS vs. 30 in the IR group).

### Differential expression analysis

iPSC lines have been demonstrated to recapitulate many Mendelian diseases, including insulin resistance resulting from severe mutations in the insulin receptor [18,19]. However, the extent to which iPSCs are a valid model for the study of common polygenic forms of IR/T2D is still unknown. To test the hypothesis that iPSCs maintain at least some differential characteristics based on the insulin sensitivity status of the individual they were derived from, we sought to analyze the differential expression (DE) signature between IR and IS iPSC lines. To that end, after alignment and feature counting, RNA-seq counts were normalized and residual expression was computed after adjusting for both technical (sequencing batch, RNA extraction method; modeled as random effects) and biological/population (reprogramming source cell, sex, ethnicity, age, BMI; modeled as fixed effects) covariates ([Figs 1 and S2](#)).

We generated several iPSC lines per individual and thus, the effective sample size of our analyses is the number of patients, not the number of samples. However, since insulin resistance status is an individual-level characteristic, we could not adjust for donor without removing the signal of interest (insulin sensitivity status). Therefore, we analyzed the data in two ways. First, we used all samples (referred to as the AS analysis) without accounting for multiple clones for each individual. Second, we averaged the residual expression levels for all clones of each individual (referred to as the average-per-patient, ApP, analysis). While generalized estimating equations, mixed models or similar techniques can be used to compute residual expression levels taking into account the correlation of samples from the same individuals, these approaches will not allow constructing co-expression and predictive networks for data with multiple clones per individual. We therefore decided to use both the AS and ApP approaches, thereby consistently using the same residual expression levels in all analyses. Importantly, iPSC lines from the same individual cluster together ([S3 Fig](#)) and samples from IR donors tend to cluster together as do samples from IS donors, whereas this is not the case according to sex ([Fig 2](#)).



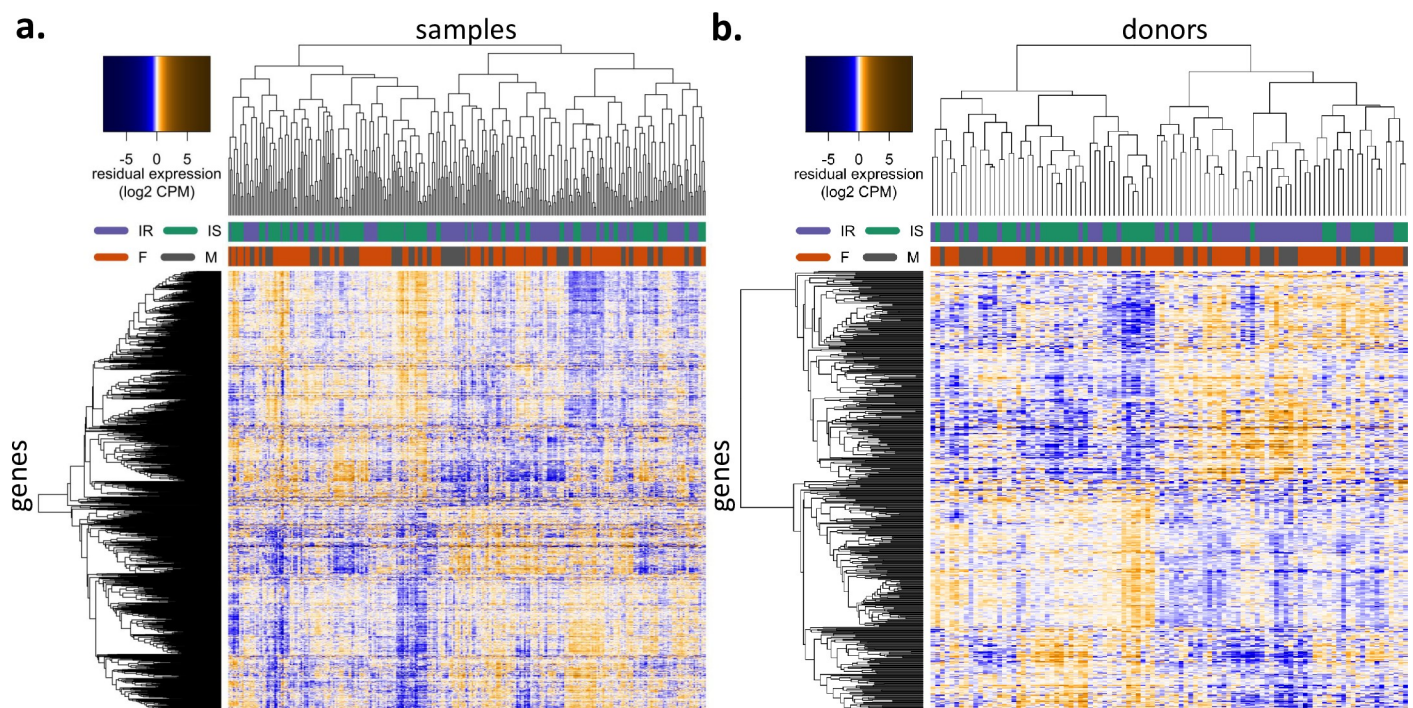
**Fig 1. Flowchart detailing the individual steps of the integrative predictive network modeling analysis pipeline and functional molecular validation: sample collection, data generation, data normalization, differential expression analysis, co-expression networks, predictive networks and key driver analysis, prioritization of KDs and molecular and functional validation.**

<https://doi.org/10.1371/journal.pcbi.1008491.g001>

We identified 1,338 differentially expressed genes between IR and IS samples in the AS analysis (FDR adjusted p-value < 0.05) whereas in the ApP analysis no significantly differentially expressed genes were identified (S2 Table). However, the comparison of the rank of the test statistics for the 500 most differentially expressed genes from both analyses shows consistent results, with very similar ranks for AS and ApP DE analysis (Spearman correlation coefficient = 0.66, median rank in both AS and ApP = 308, paired Wilcoxon signed-rank test P = 0.90) (Fig 2). This result suggests that the lack of statistically significant DE genes in the ApP analysis is due to the reduced sample size. We also performed sensitivity analyses: a) blocking by patient ID, 947 of the top 1338 most differentially expressed genes were shared with AS residuals (71% overlap) and b) using SSPG as a continuous score instead of performing an IR vs IS dichotomization showed that 840 of the top 1338 DE genes with this approach were overlapping with AS residual DE gene list (63%). As our holistic approach to compare IR and IS states was based and scaled to the entire transcriptional network, we used differential expression analysis to leverage our key driver analysis.

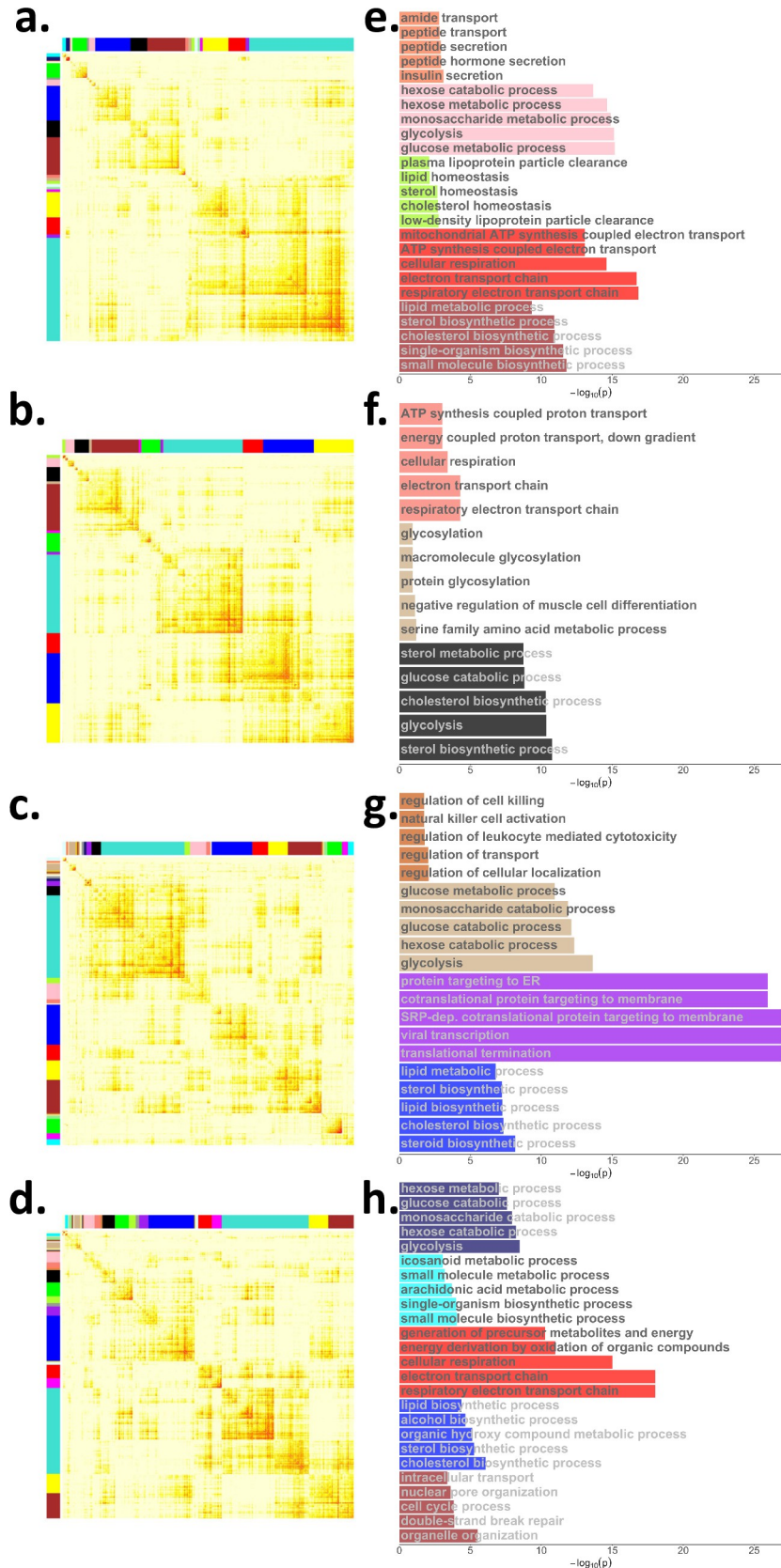
### Co-expression network analysis

We trained four co-expression networks (Fig 3A–3D) [14,20]: for each of the AS and ApP adjusted expression residuals, we built one network for IR samples and another one for IS



**Fig 2. Differential expression analysis.** (a) All-Samples (AS): 1338 differentially expressed genes with multiple testing corrected p-value < 0.05 are shown. (b) Average-per-Patient (ApP): top 500 differentially expressed genes identified in ApP are shown. Color scale indicates normalized residual expression levels (blue: low expression, orange: high expression). Columns have been arranged according to clustering of the samples / donors, (purple: insulin resistant, turquoise: insulin sensitive, orange: female, grey: male). Clustering was performed on all samples / donors agnostically of their IR / IS status and sex.

<https://doi.org/10.1371/journal.pcbi.1008491.g002>



**Fig 3. The topological overlap matrix (TOM) indicating co-expression between genes and their corresponding pathway enrichment analysis of the IR and IS co-expression network in AS and ApP.** The color bars on top and to the left of each TOM indicate different co-expression modules (WGCNA labels co-expression modules according to a sequence of colours; each colour corresponds to a different module with no further meaning of the colour labels). Only genes included in co-expression modules are shown. The color scales for the different TOMs depend on the soft power parameter used by WGCNA and hence we have not provided color scales for these heatmaps; the main point of the TOM is to show the connections (i.e. co-expression) between blocks (i.e. modules) of genes. Each module of co-expressed genes will correspond to different biological functions and thus an annotation of the modules will allow us to select modules to seed the predictive networks. (a) TOM of IR in AS; (b) TOM of IS in AS; (c) TOM of IR in ApP; (d) TOM of IS in ApP. (e)-(h) Pathway enrichment analysis results for selected modules from each co-expression networks in (a)-(d).

<https://doi.org/10.1371/journal.pcbi.1008491.g003>

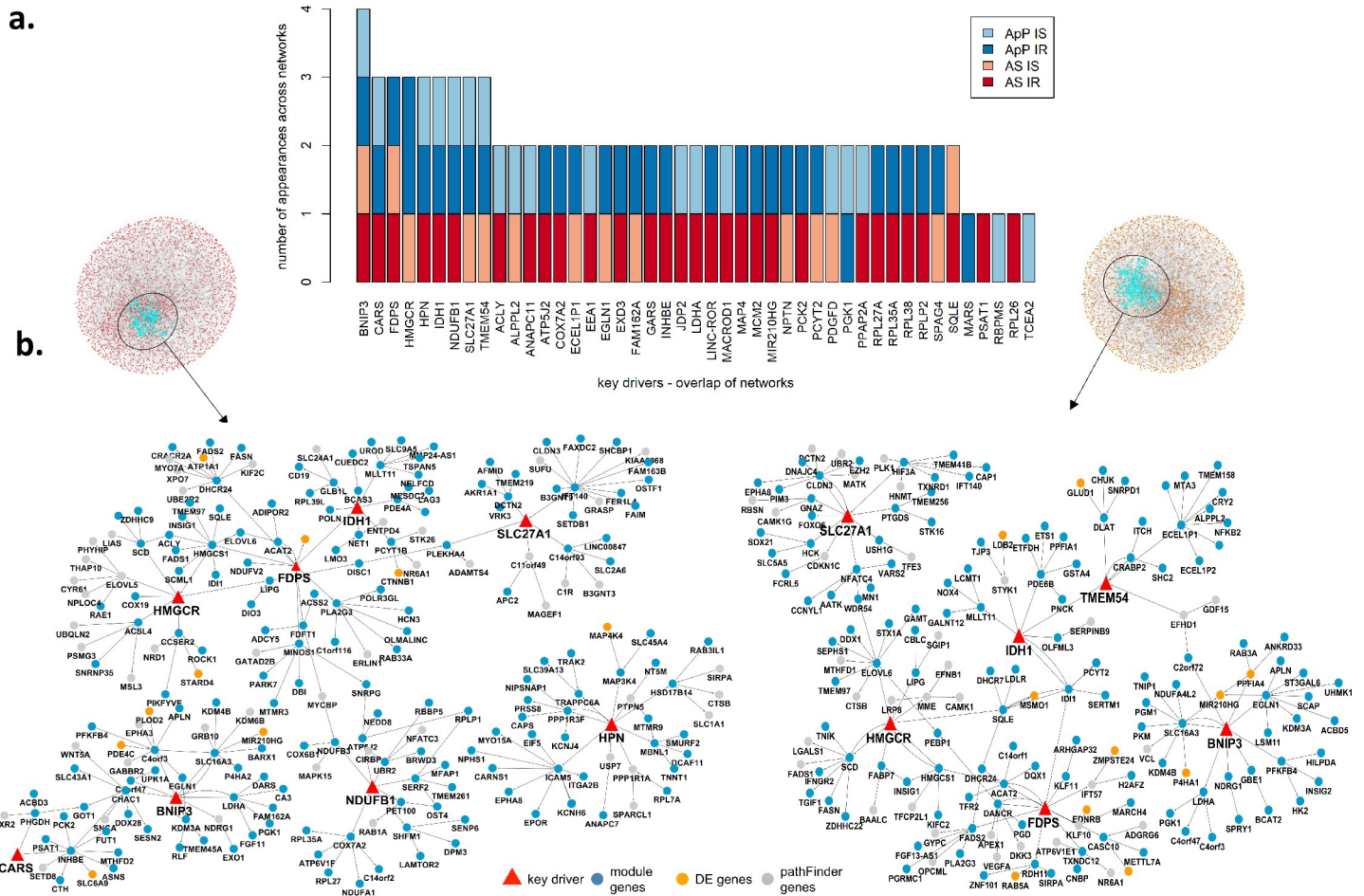
samples. Co-expression networks identify groups of genes (modules) with highly correlated expression patterns across samples, indicating that they are involved in similar biological processes. We used gene set enrichment analysis (GSEA) [21,22] and the gene ontology (GO, C5 biological processes, v5.1)[23,24] gene sets from the Molecular Signatures Database (MSigDB) [21] to test the modules of each co-expression network for enrichment in insulin or metabolism related pathways. Specifically, we identified 5 relevant modules in the AS IR network (corresponding to a total of 1,565 genes) –reflecting glucose, lipid and cholesterol metabolic processes and mitochondrial function pathways-, 3 in the AS IS network (430 genes)–glycolysis, cholesterol biosynthesis and electron transport chain-, 4 in the ApP IR network (1,689 genes)–glucose metabolic process, lipid and cholesterol associated processes and regulation of cellular localization-, and 5 in the ApP IS network (2,791 genes)–hexose metabolic process, respiratory electron transport chain, cholesterol, lipid and small molecule biosynthesis-(Fig 3E–3H). To evaluate the significance of correlations between gene pairs in the co-expression network modules, we permuted the AS and ApP residuals 10 times and computed the correlation between all pairs of genes in each module per co-expression network (AS IR, AS IS, ApP IR, ApP IS). The correlations from the permuted data formed a background distribution to calculate the p-values of the actual co-expressed gene pairs, which showed that the majority of the pairwise correlations between genes were significant in two or more modules for each of the four networks (S4 Fig).

### Predictive networks

Seeding genes for each predictive network were obtained by expanding the set of genes in the selected modules from the corresponding co-expression network by including all genes connected to any of the selected module genes in  $k = 3$  or fewer steps in a prior, cell type-specific network derived from public gene and protein interaction databases: ConsensusPathDB (CPDB) [25] and MetaCore (v6.24 from Thomson Reuters). The above seeding gene selection process ensures that genes impacting insulin sensitivity are included, while reducing the feature space by excluding irrelevant genes to train the predictive network. Our final seeding gene sets consisted of 7,250 (AS IR), 3,797 (AS IS), 8,183 (ApP IR) and 9,712 (ApP IS) genes respectively. This final seeding gene list was used in the top-down and bottom-up predictive network pipeline (see [Material and methods](#) section) to build causal network models.

### Key driver analysis (KDA) and ranking

KDA requires a starting set of genes to be analyzed and we ran the KDA multiple times, once for the genes in each co-expression module that was enriched for metabolic pathways, as well as for the DE genes from the AS and ApP analyses (5% FDR for AS and the top 500 DE genes for ApP). Consequently, for each network a given gene can be identified as a KD based on the co-expression modules and/or the DE genes, and the more often a gene is identified as KD



**Fig 4. Predictive Causal Network (PCN), key drivers and sub-networks of the tested key drivers.** (a) key driver replication across networks. For each network a given gene can be identified as a KD based on the co-expression modules and/or the DE genes, and the more often a gene is identified as KD across networks, the stronger the evidence it is implicated in the phenotype; (b) sub-networks of top 9 key drivers for IR AS (left) and IS AS (right). We highlight the key drivers (red triangles) and the 2 step downstream genes. Genes from selected modules (blue), DE genes (orange) and expanded genes through pathFinder (grey).

<https://doi.org/10.1371/journal.pcbi.1008491.g004>

across networks, the stronger the evidence it is implicated in the phenotype. Based on this approach, we performed KDA [26] on the four predictive networks and identified, a total of 231 key drivers (KD) in the AS predictive networks and 237 key drivers in the ApP networks. There were 45 key driver genes common to both sets (S4 Table). We selected the KDs for further study based on the number of appearances across networks, which rendered 9 genes (BNIP3, CARS, IDH1, NDUFB1, HMGCGR, HPN, FDPS, SLC27A1, TMEM54) that are KDs 3 or more times in the AS and ApP networks (Fig 4A and S4 Table).

For these top 9 KDs, we further looked at the topology of the local sub-networks around each of them in the four networks and represented the incoming source: selected module genes, DE genes, or expanded genes through pathFinder (Figs 4B and S5). The specific module each gene originates from and the enriched functions for those modules can be found in S6 Fig and S3 Table. In addition, we computed 2 scores (Table 1): the sum of the inverse path lengths (i.e. if a gene is k steps down/up-stream of another gene, then the inverse path length is 1/k) from each key driver to the significantly differentially expressed genes in the AS networks and the top 500 most differentially expressed genes in the ApP network (DE proximity score) and second, the difference between the sums of the inverse path lengths from each KD to other KDs downstream



**Table 1. Summary statistics and references for the top ranked key drivers.** Network appearances: number of appearances across IR and IS networks for AS and ApP DE proximity: sum of the inverse path lengths from each key driver to the significantly differentially expressed genes in the AS networks and the top 500 most differentially expressed genes in the ApP network. KD dominance: the difference between the sums of the inverse path lengths from each KD to other KDs downstream of them and the inverse path lengths to other KDs upstream of them. KD: key driver.

KD	GENE	Network appearances	DE proximity	KD dominance	REFS.
BNIP3	BCL2/Adenovirus E1B 19kDa Interacting Protein 3	4	20.30	0.00	[29–31]
FDPS	Farnesyl Diphosphate Synthase	3	16.22	1.42	none
SLC27A1	Solute Carrier Family 27 Member 1	3	10.85	0.25	[27,28]
HMGCR	3-Hydroxy-3-Methylglutaryl-CoA Reductase	3	10.15	0.58	[34,38,44–46,49]
CARS	Cysteinyl-TRNA Synthetase	3	7.80	0.00	none
TMEM54	Transmembrane Protein 54	3	7.57	1.00	none
IDH1	Isocitrate Dehydrogenase (NADP(+)) 1, Cytosolic	3	7.32	-3.25	[85–88]
HPN	Hepsin	3	6.67	0.00	none
NDUFB1	NADH:Ubiquinone Oxidoreductase Subunit B1	3	6.45	0.00	none

<https://doi.org/10.1371/journal.pcbi.1008491.t001>

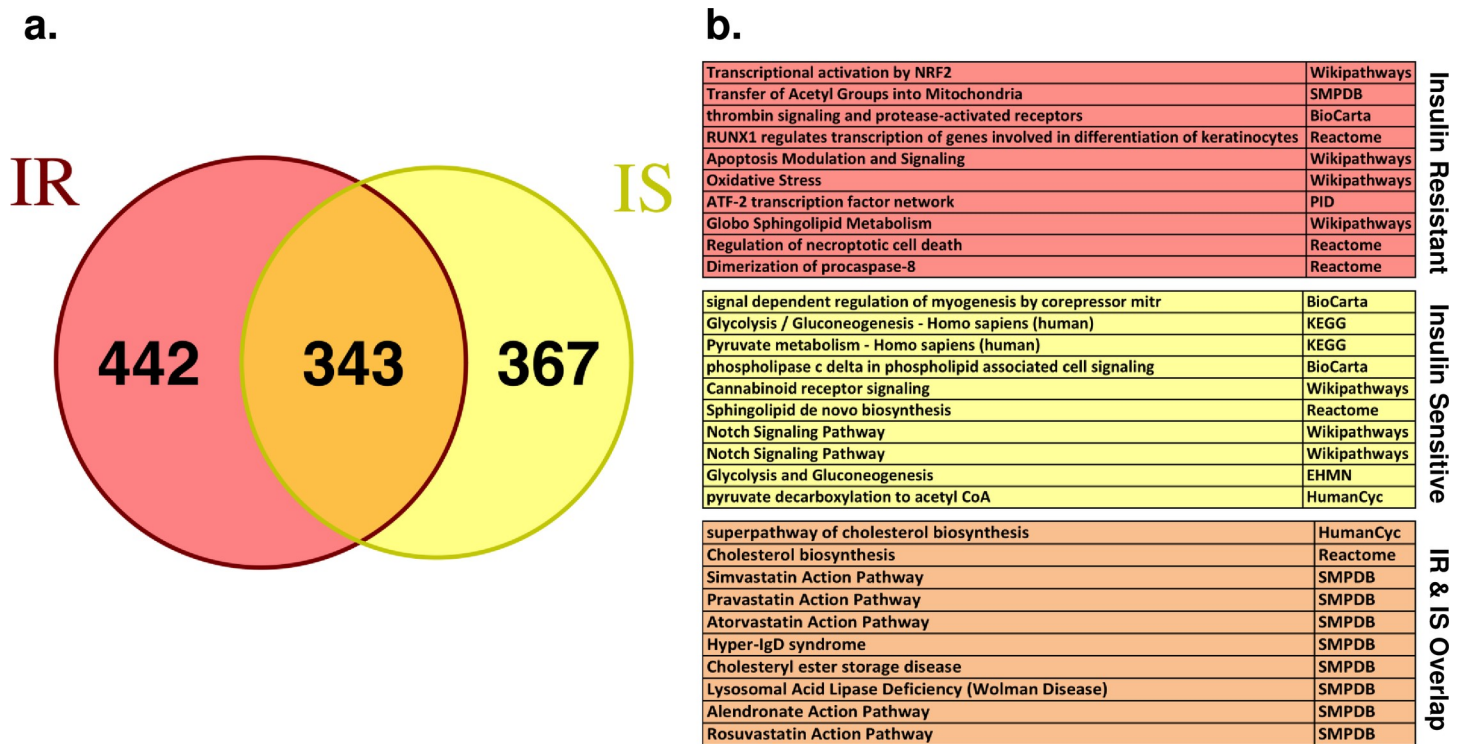
of them and the inverse path lengths to other KDs upstream of them (KD dominance score). The higher the DE proximity score, the more there are paths from this KD to DE genes and/or the shorter these paths are; the higher the KD dominance score, the more other KDs are more directly downstream of this KD and/or the fewer other KDs are directly upstream of it.

Finally, we performed a directed literature search combining the different KDs with the terms insulin, glucose, and diabetes. The results of the relevant publications associating the KDs to insulin sensitivity are summarized in **Table 1**. Briefly, both *BNIP3* and *SLC27A1* have been strongly associated with insulin resistance phenotypes. Solute Carrier Family 27 Member 1 (*SLC27A1*, also known as *FATP1*) is an insulin-sensitive fatty acid transporter involved in diet-induced obesity and has been associated to IR in skeletal muscle [27,28] and *BCL2/Adenovirus E1B 19kDa Interacting Protein 3* (*BNIP3*) is essential for mitochondrial bioenergetics during adipocyte remodeling [29], regulates mitochondrial function and lipid metabolism in the liver [30] and in conjunction with *PPAR $\gamma$*  couples mitochondrial fusion-fission balance to systemic insulin sensitivity [31]. Although informative about the quality of our KDA, the body of publications related to these two KDs decreased the interest on further validation. Among the remaining top KDs, *HMGCR* and *FDPS* have the highest combined DE proximity and KD dominance scores and both genes participate, together with squalene epoxidase (*SQLE*)-another KD that appears in both AS and ApP networks (**S4 Table**)- in the cholesterol biosynthesis pathway. Given, i) the high DE proximity and KD dominance scores, ii) the shared metabolic pathway, iii) the widespread use of statins (*HMGCR* inhibitors) as therapeutic drugs to lower cholesterol levels in patients with high LDL-cholesterol, and iv) the emerging role of *HMGCR* in energy balance, metabolism and diabetes risk [32–35], we sought to validate our KDA, both transcriptionally and functionally, focusing on *HMGCR*, *FDPS* and *SQLE*.

## Network validation

We validated the causal IR/IS networks and the key driver analysis using the DE gene signature from an *HMGCR* inhibition experiment in iPSC cell lines derived from both IR ( $n = 6$ ) and IS individuals ( $n = 6$ ). For each iPSC line in this experiment (**S5 Table**) we generated RNA-seq data in presence or absence of atorvastatin, an *HMGCR* inhibitor (statin) widely used in patients with hypercholesterolemia [36]. Previous efforts have validated predictive networks through similar approaches [12,37]

The comparison of atorvastatin-treated and untreated samples resulted in a list of 3205 DE genes (**S6 Table**) that showed the highest enrichment for statin action pathway and cholesterol



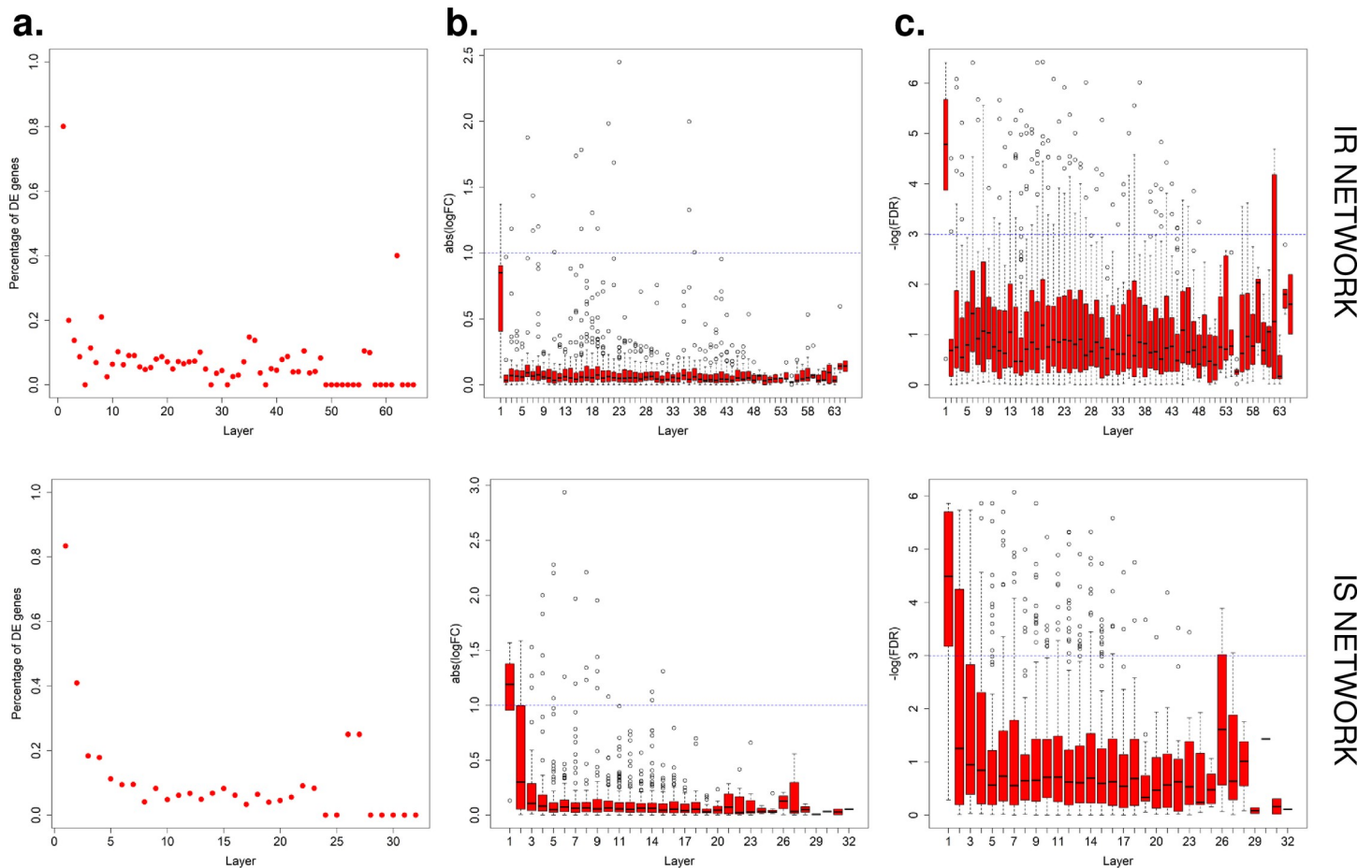
**Fig 5. Differential pathway enrichment after HMGCR inhibition in IR vs IS iPSCs.** (a) Venn diagram for the IR-specific (442 genes), IS-specific (367 genes) and IR/IS overlapping (343 genes) DE genes due to HMGCR inhibition. (b) Pathway enrichments for the groups defined in (a). Top-10 significant pathways are shown for each group (full list of pathway enrichment can be found in [S7 Table](#)).

<https://doi.org/10.1371/journal.pcbi.1008491.g005>

biosynthetic pathway and other related pathways ([S7 Table](#)), suggesting that HMGCR inhibition triggers a transcriptional compensatory response to balance the decrease in the cholesterol pathway intermediates.

Next, we wanted to compare the specific responses of IR vs. IS iPSCs to atorvastatin treatment. When considering IR or IS groups independently, the number of DE genes was reduced to 785 (IR) and 711 (IS) ([S6 Table](#)). Surprisingly the DE genes between IR and IS samples are just partially overlapping (343/785 or 343/711, respectively) ([Fig 5A](#)) which suggests a differential response to atorvastatin treatment based on the IR/IS status of the donors. Moreover, pathway enrichment analysis ([S7 Table](#)) for the 442 IR-specific and 367 IS-specific DE genes demonstrated striking differences in the enriched pathways between IR and IS iPSCs ([Fig 5B](#)), which could be related to the disproportionate incidence of T2D in insulin resistant patients under statin treatment [38]. Although atorvastatin treatment translates the perturbation of a metabolic pathway into measurable transcriptional changes and gives clues about HMGCR functionality and its association to insulin sensitivity, it requires of novel additional analyses to validate the predictive network.

To validate our network structure, i) we calculated the percentage of genes in each downstream layer of HMGCR in the network that are significantly altered (FDR<0.05) in gene expression in HMGCR inhibition experiment. We found that for both IR and IS networks more than 80% of the genes in the first layer downstream of HMGCR are DE genes and that this percentage decreases as the distance to HMGCR increases in the network ([Fig 6A](#)); ii) we also compared DE gene fold-changes (log FC) ([Fig 6B](#)) and associated significance (-logFDR) ([Fig 6C](#)) to the AS network topology. Among the genes located downstream of HMGCR in both the IR and IS networks, the fold change and their significance decrease as the distance to



**Fig 6. Predictive Network Validation (AS network) by RNA-seq data in Atorvastatin treated iPSC lines.** (a) The percentage of DE genes decreases as the distance (layers) from HMGCR increases. (b) Among the genes located downstream of HMGCR, the fold change decreases as the distance to HMGCR increases. (c) The p-value of differential expression analysis from Atorvastatin experiment decreases along the steps down from HMGCR. Upper lane: AS IR network. Lower lane: AS IS network.

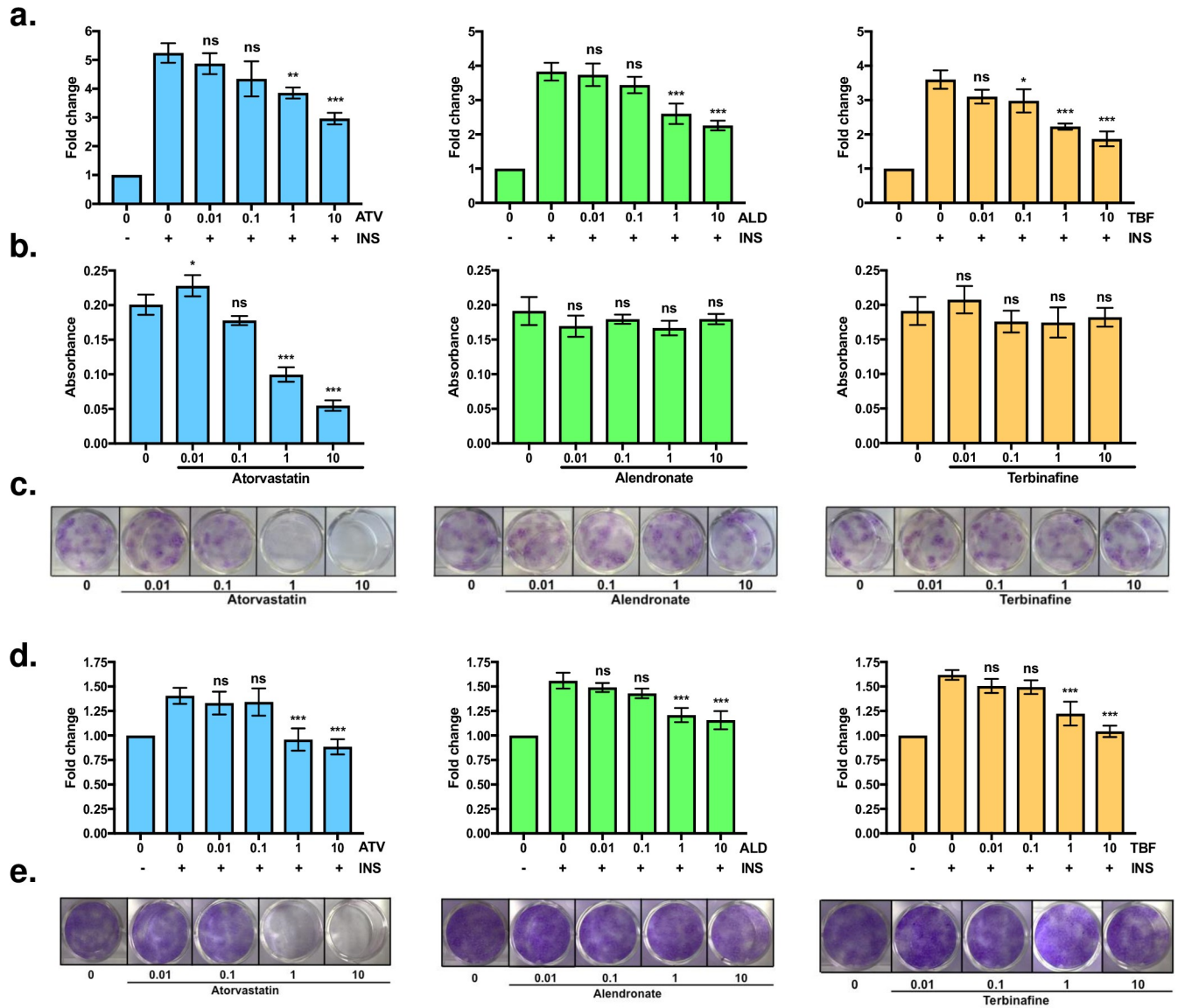
<https://doi.org/10.1371/journal.pcbi.1008491.g006>

HMGCR increases (Fig 6B and 6C). The same pattern was observed in the ApP network (S7 Fig). Finally, all four predictive networks (AS IR, AS IS, ApP IR and ApP IS) show a significant enrichment downstream of HMGCR for DE genes due to HMGCR inhibition (S8 Table).

Thus, our results confirmed that percentage of DE genes, DE gene fold change and associated significance decreases as the distance (number of layer/steps away) from the perturbed target increases, and that DE genes are enriched in the downstream steps of HMGCR compared to non-DE genes. Taken together, the HMGCR-inhibition experiment validated the predictive networks and their topological structure.

### Functional validation

We next sought to functionally validate the prioritized KDs -*HMGCR*, *FDPS* and *SQLE*- in processes associated with insulin sensitivity and in particular, to insulin mediated glucose uptake in relevant metabolic cell types such as adipocytes and SKMCs. To that end, we differentiated the Simpson-Golabi-Behmel syndrome (SGBS) human preadipocyte line [39] and the human SKMC line HMCL-7304 [40] to terminally differentiated adipocytes and myotubes, respectively. We focused our validation efforts on *HMGCR*, *FDPS* and *SQLE* as these three key drivers participate in the same metabolic pathway-cholesterol biosynthesis- and, in addition,



**Fig 7. Functional validation of key driver genes.** (a) Insulin mediated glucose uptake assay in mature human adipocytes. Fold change values are shown with respect to no insulin (-). Each panel shows the effect of varying concentrations of inhibitor for HMGCR (atorvastatin), FDPS (alendronate) and SQLE (terbinafine). (b) Adipogenic differentiation assay. The effect of different concentrations of the inhibitors over SGBS adipogenesis is measured by absorbance measurement of Oil-O-Red emission at 500nm. (c) Growth assay in human SGBS preadipocytes. Crystal violet staining after 12 days of assay in presence/absence of the inhibitors. (d) Insulin mediated glucose uptake assay in mature human SKMCs. See (a). (e) Growth assay in human SKMCs. See (b). Results represent mean±SD, and statistical significance was evaluated through One way ANOVA \*p<0.05, \*\*p<0.01, \*\*\*p<0.001 compared to insulin condition without inhibitors (a and d) or basal differentiation (c).

<https://doi.org/10.1371/journal.pcbi.1008491.g007>

statins (*HMGCR* inhibitors] are at the center of an intense debate about their effect on insulin resistance and type II diabetes risk [33,41–45].

To functionally inhibit the three candidate genes, we used well-described and widely used chemical inhibitors: atorvastatin (targeting *HMGCR*), alendronate (for *FDPS*) and terbinafine (for *SQLE*) (S8 Fig). As shown in Fig 7A, all three inhibitors decrease insulin mediated glucose uptake in human adipocytes. However, only *HMGCR* inhibition demonstrates a detectable decrease of preadipocyte growth and differentiation efficiency (Fig 7B and 7C). Along the same lines, atorvastatin inhibits both insulin mediated glucose uptake and cell proliferation in

SKMCs, while alendronate and terbinafine show only a significant effect on glucose uptake (Fig 7D and 7E).

Our results suggest that data driven co-expression and predictive networks combined with key driver analyses are powerful tools for the discovery of novel genes involved in IR.

## Discussion

Although GWAS studies have targeted T2D and insulin resistance-associated glycemic traits, the success in identifying new genes that contribute to insulin resistance risk has been fairly limited. Our group has previously demonstrated that GWAS studies based on gold-standard measurements allows the discovery of novel genes associated with insulin resistance [6] but the power to detect novel loci associated to IR has been limited by sample size. In addition, the genetic complexity and multicellular targets of insulin resistance advocate for the development of new cellular systems and holistic genetic approaches.

With this goal in mind, we generated an iPSC library with accurate measurements of insulin sensitivity that reflects the broad spectrum of insulin responses in human populations. Given our limited sample size (52 IR vs. 48 IS individuals for a total of approximately 300 iPSC lines), and to overcome the limitations of the traditional DE analyses, we developed a more holistic view of the genetic networks associated to IR through the construction of co-expression networks for both IR and IS iPSC lines. In addition, for network construction we performed a dual approach where we used the gene expression values for all samples (AS) (to increase power) or the average per patient (ApP) (to increase stringency). One limitation of our study is that given the complex polygenic nature of insulin sensitivity and given the complexity of the generated data and the multiple sources of variance to gene expression, the multiple corrections that were needed to perform the DE analyses could potentially mask some of the genetic signals associated to insulin sensitivity. Moreover, the extent to which different DE analyses, like using SSPG as a continuous variable instead of dichotomizing IR and IS status, would affect network construction has not been explored in the present study. However, the constructed networks highlighted co-expression modules enriched for cellular functions like respiratory electron transport chain, glycolysis, cholesterol and steroid biosynthesis and glucose metabolism that are intimately associated with insulin sensitivity-associated processes. We performed predictive network and key driver analyses to investigate the central genetic nodes that control the aforementioned modules and functions and thus, are most likely to be involved in the etiology of insulin resistance.

To better delimit and rank our key driver list we considered only the KDs defined as such in both AS and ApP approaches (45 genes) and then we considered the total number of appearances in the 4 constructed networks (AS IR, AS IS, ApP IR and ApP IS), which identified 9 top key drivers. As highlighted in Table 1, *IDH1*, *BNIP3* and *SLC27A1 (FATP-1)* have been shown to participate in functions associated with insulin sensitivity or have been directly associated to insulin resistance or type 2 diabetes. Among the rest of the selected key drivers, the top two key drivers with the highest DE path and KD path values, which represents the connectivity of a given KD to DE genes and to other KDs, are Farnesyl Diphosphate Synthase (*FDPS*) and 3-Hydroxy-3-Methylglutaryl-CoA Reductase (*HMGCR*) which coordinately participate in the cholesterol biosynthetic pathway. Moreover, another gene participating in this pathway, *SQLE* is among the 45 KDs shared between both approaches. Meta-analysis of clinical trials with statins (*HMGCR* inhibitors) have shown an increase in T2D incidence [44–46], which seems to be dose related [35,43] and that affects insulin resistant individuals in a disproportionate way [38]. In addition, alleles in *HMGCR* that lower LDL-C confer an increased risk of developing T2D and individuals with familial hypercholesterolemia are protected against

T2D [34,47], leading to speculation that statins affect insulin sensitivity and insulin secretion [45,48], although the exact cellular and molecular mechanisms to such an increase in T2D risk are still not well understood.

The predictive network not only illustrates co-regulated genes in the same pathway but can also demonstrate causality upstream and downstream of a given gene. There have been successful efforts to validate predictive networks [12,37], and therefore, it was critical to show that we could capture the downstream effector genes of key drivers in the predictive networks. Our empirical network validation through HMGCR inhibition demonstrated enrichment for DE genes and log fold change in the downstream proximity of HMGCR, which validates the overall structure of the network. In addition, we observed a differential response to atorvastatin treatment depending the IR or IS state of the iPSC tested, which could be related to the increased incidence of T2D under statin treatment that seems to specially affect IR individuals [34,38,41].

Our functional validation assays in human preadipocyte (SGBS) cells demonstrate that HMGCR inhibition decreases preadipocyte proliferation, and differentiation and insulin mediated glucose uptake in mature adipocytes, similar to previous results in mouse preadipocyte lines [49–51]. In addition, both proliferation and insulin mediated glucose uptake are affected in the human SKMCs (HMCL-7304), treated with atorvastatin. FDPS and SQLE inhibition have also a significant effect decreasing insulin mediated glucose uptake in human adipocytes and myotubes, while not affecting proliferation or differentiation. These results suggest that the effects on proliferation and differentiation of adipocytes and SKMCs are independent of the disturbance in insulin mediated glucose uptake. In addition, *SQLE* inhibition exerts a comparable effect on insulin mediated glucose uptake when compared to *HMGCR* inhibition, which suggests that the underlying mechanisms could be mediated by the deregulation of intracellular or membrane bound cholesterol levels [32,52]. Other proposed mechanisms to affect insulin facilitated glucose uptake include decreased expression levels of *SLC2A4* (*GLUT-4*) and caveolin-1 [51,53], disturbed *RHOA* and *RAB4* signaling lowering *GLUT-4* levels in the plasma membrane [54,55], perturbation of the insulin signaling pathway [55,56] and accumulation of free fatty acids [49,57]. Although our results suggest that the cholesterol biosynthetic pathway is directly involved in the insulin-mediated glucose uptake impairment induced by atorvastatin, it is highly probable that additional mechanisms are participating. In addition, different types of statins could exert differential effects in a concentration, species (mouse vs. human) or cell specific context (adipocytes vs. skeletal muscle cells). In any case, further studies are necessary to dissect the specific mechanisms by which statins increase the risk of developing T2D.

In summary, our work suggests that: i) iPSC retain a donor-specific signature [9], ii) co-expression and predictive networks combined with key driver analyses uncovered robust candidates to participate in IR, iii) IR iPSCs have a differential response to *HMGCR* inhibition when compared with IS cells and iv) the cholesterol biosynthetic pathway is involved in the insulin-mediated glucose uptake impairment observed in human adipocytes and SKMCs. Taken together, these results suggest that iPSCs technology will offer a novel and sophisticated model for the study of IR and the associated cardiovascular disease, especially when relevant metabolic (adipocytes or SKMCs) and vascular (endothelium) cell types are generated from our iPSC library with accurate measurements of insulin sensitivity.

## Material and methods

### Ethics statement

The study included 201 subjects who had volunteered for prior studies between October 2002 and October 2013 and were in general good health. The individuals were re-contacted and re-

consented for this study. Stanford Institutional Review Board approved the study protocol, and all subjects gave written informed consent for study participation.

### Patient recruitment, biological parameters and insulin sensitivity measurement

Patient recruitment and blood sampling were performed as previously described [9]. Insulin sensitivity measurement was performed by a modified insulin suppression test in accordance with Knowles et al. [7] See [S1 Table](#) for complete demographic data.

### RNA-Seq processing

STAR v2.4.0g1 [58] was used to align RNA-seq reads to the human genome built GRCh37. Using featureCounts v1.4.4 [59], we counted the uniquely mapping reads overlapping genes as annotated by ENSEMBL v70.

### Statistical analyses and data processing

Unless otherwise specified, statistical analyses and data processing steps were done using R v3.0.3 [60]. Code for the various analyses described below has been published on GitHub (<https://github.com/gitMarch/iPSCsInsulinResistance> and <https://github.com/zhukuixi/KDA>).

### Expression data normalization and covariate adjustment

After filtering out lowly expressed genes (at least 1 counts-per-million (CPM) in 30% or more of samples), we have data for 15,294 genes left for analysis. We used the R packages edgeR [61] and limma [62] to normalize the RNA-seq expression data. The edgeR function *calcNormFactors()* is used to compute TMM [63] weights which are applied using the limma function *voom()*. The final normalized expression values that are output by *voom()* are on the log<sub>2</sub> counts-per-million (CPM) scale.

All RNA-seq data analyses were performed on expression residuals corrected for the effects of technical (sequencing batches and RNA preparation kits, reprogramming source cell) and patient covariates (sex, ethnicity, age, BMI) in line with existing differential expression analysis literature [64–66]. Batch and RNA preparation kit were adjusted for as random effects using the variancePartition R library [67] whereas reprogramming source cell, and patient characteristics were adjusted for as fixed effects using the limma [60] package. Each of the variables we adjusted for corresponds to technical or biological variation that will mask independent genetic signals of insulin sensitivity. At the same time each of these variables impacts the expression levels of a significant number of genes, and hence, any differences in these factors between the IR/IS groups are likely to confound the biological variation due to IR. The variance partition analysis in [S2 Fig](#) provides insights into the gene expression variability correlated with these covariates.

As detailed in the results section, due to having multiple clones per patient, we computed two sets of expression residuals: one (AS) using all samples from all clones for every patient and one (ApP) where we averaged residuals per patient.

### Genomic data and expression quantitative trait (eQTL) analysis

We have described these processing and analysis steps previously [9]: “Genotype data were filtered to remove markers with over 5% missing entries, minor allele frequency below 1% and Hardy-Weinberg p value  $< 10^{-6}$ . Genotypes were phased with SHAPEIT v2.r790 [67], and missing genotypes were imputed with Impute2 v2.3.2 [68] using the reference panel from the

1000 Genomes Project Phase 3 [69]. Markers with high imputation quality ( $\text{INFO} > 0.5$ ; [68] and minor allele frequency over 1% were retained for downstream analyses.

Following standard practice, only individuals of European ancestry were included in the eQTL analysis in order to avoid false positives due to the correlation between ancestry and gene expression. Principal components analysis based on genome-wide genotype data identified 81 individuals of European ancestry for eQTL analysis. eQTL analysis was performed with MatrixEQTL v2.1.1 [70] using the first 5 genotype principal components as covariates. Latent variables were identified in the gene expression data using PEER v1.0 [71]. [ . . . ]. Cis-eQTL analysis considered markers within 1Mb of the transcription start site of each gene. False discovery rates were computed following Benjamini–Hochberg.” The purpose of the cis-eQTL analysis in this work was to have structural priors for the predictive networks (see “Predictive networks” section below).

### Differential expression analysis

For the expression residuals used for the co-expression and predictive network analyses we adopted 2 analysis streams: one where we use all samples without adjusting for patient ID (AS) and one where we average residuals per patient (ApP). Differentially expressed genes were determined using a linear model, as implemented in the `lmFit` function from the `limma` package v3.18.13 in R. Statistical significance was assessed using a cut-off of 0.05 on FDR adjusted p-values.

### Co-expression networks and selection of co-expression modules

We constructed co-expression networks using the `coexpp` R package [20], which provides an optimized workflow for the `WGCNA` [14] R package (v1.14–1 used here together with R v3.0.3) for large numbers of genes. The soft threshold powers used for the AS IR, IS and ApP IR, IS co-expression networks were 7.5, 9.0, 7.5 and 9.5 respectively. These were chosen by the `pickSoftThreshold()` function from the `WGCNA` package with an  $R^2$  cut-off of 0.8, unless visual inspection of the  $R^2$  curve as a function of power revealed a plateau near 0.8 and a lower power achieved an  $R^2$  of almost 0.8 (e.g. between 0.78 and 0.8). In this case, that lower power was chosen instead. All 15,294 genes that passed the low-expression threshold were used to build the co-expression networks. Seeding genes for the predictive network (specifically, the input to `pathFinder`) were selected to be the genes in co-expression network modules statistically enriched ( $\text{FDR} < 0.05$ ) for GO terms relevant to insulin resistance related traits (biological processes only).

### Predictive networks

Bayesian networks (BN), which provides a natural framework for capturing causality among highly dissimilar types of data, stochastic processes of biology systems, and noise, have become increasingly popular for modeling biology systems [72–75] due to their inherent capability to integrate multi-scale ‘omics’ data such as genetic context, transcriptomics, proteomics, metabolomics, epigenomics and literature knowledge. While Bayesian networks are useful in deciphering causality of molecular interactions, one fundamental problem is that BN can NOT infer causality among statistically equivalent structures, i.e. multiple network structures with opposite causality direction fit equally well to the data. The problem of partial causality in BN has limited their effectiveness in identifying high-quality key drivers by inducing randomness to the causality of learned structure.

Recently, we developed the top-down and bottom-up predictive network pipeline to build causal predictive network models [76], which leverages the bottom-up belief propagation



engine [77,78] as a sub-routine to infer causality among equivalent structures. Biochemical reactions can be modeled as nonlinear perturb-response relations, e.g. hyperbolic response, sigmoidal graded/switched response, rise/fall pulsed response, complex non-monotonic or periodical oscillation [79]. The bottom-up method leverages the nonlinearity of biochemical reactions to infer causality of a molecular interaction, i.e. fitting better to the data along true causal direction than false causal direction, therefore, breaks the statistical equivalence. By integrating the novel bottom-up causality inference approach with (top-down) Bayesian networks, the integrated top-down & bottom-up predictive network platform will result in complete causal network with discerned causality among equivalent structures. This pipeline inherits the advantage of BN in integrating the multi-scale 'omics-' data (genetics, genomics, proteomics, metabolomics, epigenomics) to construct multi-scale network models. Consequently, the nodes in a predictive network can represent any variable of interest within the biological system, e.g. levels of gene expression [9,76,80] the genotype of a locus [9], the activity of a protein [76], or the abundance of a metabolite, to name just a few. In addition, the predictive network inherited the capability to predict new molecular phenotype upon genetic perturbations through its internal bottom-up belief propagation engine [77,78]. Also, the genotype data is incorporated as *cis*-eQTL genes in the model where they are constrained to be the top node (without other parents). As *cis*-eQTLs causally affect the expression levels of neighboring genes, they can serve as a source of systematic perturbation to infer causal relationships among genes [9,81,82]. Consequently, we incorporated *cis*-eQTL genes into each network as structural priors. We used the genes in the selected modules from co-expression networks as seeding genes for predictive network modeling.

To build a predictive network, our platform involves the following steps: i) Evaluate the variance of expression level attributed by every technical and clinical covariate using variance partition to adjust a subset of covariates. ii) Perform differential expression (DE) analysis to derive significant DE genes/proteins/metabolites between groups of interest. iii) Build WGCNA co-expression networks to select a set of modules enriched for significant functions. iv) Extract a list of seeding genes from the selected modules and expand this seeding gene set by prior cell type-specific signaling pathways built separately through pathFinder.

### Key driver analysis (KDA) and prioritization of top hits

The fundamental idea of KDA algorithm is to search for master regulators upstream of a set of user-defined downstream effector genes (in our study, we used the genes from the selected modules and the DE genes) given a causal network structure based on Fisher's exact test. KDA, was performed using a modified version of the R package KDA [26] that we have published in <https://github.com/zhukuixi/KDA>. Specifically, first a background sub-network is defined by KDA by looking for a K-step upstream neighborhood round each node in the target gene list in the network. Second, starting from each node in this sub-network, KDA evaluates the enrichment of downstream neighborhoods (for each step size from 1 to K) for the target gene list. K = 6 was used in this paper. This cutoff was chosen based on practical considerations (K will depend on the size and complexity of the network) and while we can use other cut-offs, K = 6 is commonly used in previous studies and is the default value suggested in the KDA software [9,26,83,84]. This K value is preferred because the key drivers within K steps can all be detected. However, as we learned from our knockdown experiment and Fig 6, when K gets too large, the impact of a gene on the downstream genes diminishes exponentially. Therefore, we believe that K = 6 is a reasonable choice for this study. Finally, we took the overlap of key drivers identified from the AS and ApP networks and further ranked the top 9 KDs (described in the results section and on Table 1).

### pathFinder, a fast graphical algorithm

pathFinder is a graphical algorithm we developed and described in [9]. The purpose of pathFinder is to extract neighborhood structures, specifically to expand an initial set from a larger background network. In particular for this paper we expanded the gene set arising from the selected modules.

### HMGCR inhibition in iPSC lines

iPSCs from 6 IS and 6 IR individuals were used in this validation experiment. These 6 individuals were chosen based on the extremes of the distribution for insulin sensitivity and to reduce the confounding genetic noise that intermediate samples could generate. We selected individuals with an SSPG over 200 for the IR group and individuals with an SSPG below 100 for the IS group). In addition, we selected these samples in order to be able to match gender, age and BMI between IR and IS samples. Full covariate information for these individuals can be found in [S5 Table](#). The cells were maintained in feeder-free conditions using mTesr1 (Stem Cell technologies, Inc) supplemented with 1 mM L-Glutamine, 1mM Penicillin-Streptomycin, 0.1 µg/ml Fungizone.) on 5% matrigel coated 6-well TC plates. For passaging, cells were washed once with PBS and treated with pre-warmed 1 mM EDTA (Sigma), incubated at 37 degrees for 1–5 minutes, and resuspended in fresh mTesr1 medium 2 µM Thiazovivin (Millipore). After 12 hours incubation in Thiazovivin, medium was changed daily with fresh mTesr1. Cells were grown to ~90–100% confluency, washed once with PBS and were treated with either DMSO (D) or 1 µM Atorvastatin (A) (Selleckhem) for 12h. Cells were washed once in PBS and harvested for RNA extraction using PureLink RNA mini kit (Thermo Fisher Scientific). Total RNA was quantified using a Nanodrop (Thermo Scientific). RNA samples with a A260/280 ratio <1.8 or >2.3 were excluded from further processing and the RNA was sequenced using the Illumina HiSeq 2500 system.

### Cell culture

Simpson-Golabi-Behmel syndrome (SGBS) cells (human preadipocytes) were provided by Dr. Martin Wabitsch (Ulm University, Ulm, Germany). SGBS cells were cultured in DMEM/F12 supplemented with 10% FBS, 33µM biotin, and 17µM panthotenate. The SKMC line HMCL-7304 was provided by Institute of Child Health (ICH), University College London. Cells were cultured in SKMC growth medium (PromoCell). iPSCs were generated and cultured as previously described [9].

### Adipogenic and skeletal muscle differentiation

For adipogenic differentiation SGBS cells were grown to confluency and subjected to a two-step differentiation process. Cells were first exposed for 3 days to media composed of DMEM/F12 supplemented with 0.01mg/mL of transferrin, 20µM of insulin, 100nM cortisol, 0.2nM 3,3',5-Triiodo-L-thyronine, 25nM dexamethasone, 250µM 3-Isobutyl-1-methylxanthine, and 2µM rosiglitazone. Afterwards, cells were exposed to DMEM/F12 supplemented with 0.01mg/mL of transferrin, 20µM of insulin, 100nM cortisol, and 0.2nM 3,3',5-Triiodo-L-thyronine for additional 12 days. HMCL-7304 cells were differentiated in presence of SKMC differentiation medium (PromoCell) for 4–5 days before glucose uptake was performed.

For quantification of the effect on adipogenic differentiation, chemical inhibitors for HMGCR (atorvastatin), FDPS (alendronate), and SQLE (terbinafine) were added at day 0 of differentiation at different concentrations (10nM, 100nM, 1µM, 10µM.). Differentiation quantification was performed with Oil Red O. Differentiated SGBS cells were fixed with 10%

formalin for 10 minutes at room temperature. After washing with 60% isopropanol, samples were incubated in Oil Red O (Sigma-Aldrich) for 10 minutes at room temperature. Oil Red O was eluted with 100% isopropanol for 10 minutes. The solution was then transferred into a 96-well plate and absorbance measured at 500nm.

### Glucose uptake

Differentiated SGBS or HMCL-7304 cells were pretreated for 24 hours with different concentrations (10nM, 100nM, 1uM, 10uM) of the chemical inhibitors for HMGCR (atorvastatin), FDPS (alendronate), and SQLE (terbinafine). After the preincubation, the cells were starved for 2 hours prior to 30 minute of 100nM insulin stimulation at 37 degrees. Following stimulation, cells were incubated with Krebs-Ringer bicarbonate-HEPES (KRBH) buffer (130mM NaCl, 5mM KCl, 1.3mM CaCl<sub>2</sub>, 1.3 mM MgSO<sub>4</sub>, 25mM HEPES, pH 7.4) containing 100uM 2-deoxy-D-glucose, and 1uCi/ml 2-deoxy-D-[1,2-<sup>3</sup>H]glucose for 10 minutes at room temperature. The cells were then washed with PBS, harvested in 300uL of M-PER lysis buffer (ThermoFisher), and then added to scintillation vials containing 4.75mL of scintillation fluid (Perk Elmer). Radioactive counts were determined with a scintillation counter (Model ID: Beckman LS6500). Excess samples were subjected to BCA assay for protein quantification and normalization of radioactive counts. All samples were represented as fold change compared to the unstimulated (no insulin) condition.

### Growth assay

SGBS or HMCL-7304 cells were plated at 50 or 100 cells/cm<sup>2</sup> in 12 well plates and were grown for 12 to 14 days in the absence or presence of 10nM, 100nM, 1uM, 10uM of atorvastatin, terbinafine or alendronate. After the treatment, the cells were fixed in cold methanol for 15 minutes and stained with crystal violet for 10 minutes. Dye excess was washed with water and pictures taken immediately afterwards.

## Supporting information

### S1 Fig. Reprogramming efficiency and SSPG distribution.

(TIF)

### S2 Fig. Variance partition.

(TIF)

### S3 Fig. Hierarchical clustering of samples.

(TIF)

### S4 Fig. Significance of gene-pair correlation in the four co-expression networks.

(TIF)

### S5 Fig. ApP subnetworks for the top 9 KDs.

(TIF)

### S6 Fig. Module annotation for the genes in the four subnetworks.

(TIF)

### S7 Fig. AS network validation.

(TIF)

### S8 Fig. Cholesterol biosynthetic pathway and location of HMGCR, FDPS and SQLE.

(TIF)

**S1 Table. Patient biometric information.**

(XLSX)

**S2 Table. IR vs IS DE genes and rank comparison.**

(XLS)

**S3 Table. Go term enrichment for co-expression modules.**

(XLSX)

**S4 Table. Key driver analysis.**

(XLSX)

**S5 Table. Patient information atorvastatin experiment.**

(XLS)

**S6 Table. ATV experiment DE genes.**

(XLSX)

**S7 Table. Pathway enrichment atorvastatin experiment.**

(XLS)

**S8 Table. HMGCR inhibition DE genes enrichment in downstream of HMGCR in predictive networks.**

(XLSX)

## Acknowledgments

We acknowledge Institute of Child Health (ICH), University College London for providing the HMCL-7304 immortalized SKMC line. We want to thank Johanne M Justensen for helping with the upload of the sequencing data to dbGAP.

## Author Contributions

**Conceptualization:** Ivan Carcamo-Orive, Marc Y. R. Henrion, Eric E. Schadt, Thomas Quertermous, Joshua W. Knowles, Rui Chang.

**Data curation:** Ivan Carcamo-Orive, Marc Y. R. Henrion.

**Formal analysis:** Ivan Carcamo-Orive, Marc Y. R. Henrion, Kuixi Zhu, Noam D. Beckmann.

**Funding acquisition:** Eric E. Schadt, Thomas Quertermous, Joshua W. Knowles, Rui Chang.

**Investigation:** Ivan Carcamo-Orive, Paige Cundiff, Sunita L. D'Souza.

**Methodology:** Ivan Carcamo-Orive, Marc Y. R. Henrion, Kuixi Zhu, Martin Wabitsch, Rui Chang.

**Project administration:** Ivan Carcamo-Orive, Joshua W. Knowles.

**Resources:** Thomas Quertermous, Joshua W. Knowles, Rui Chang.

**Software:** Marc Y. R. Henrion, Kuixi Zhu, Noam D. Beckmann, Sara Moein, Zenan Zhang, Melissa Alamprese, Rui Chang.

**Supervision:** Eric E. Schadt, Thomas Quertermous, Joshua W. Knowles, Rui Chang.

**Validation:** Ivan Carcamo-Orive, Marc Y. R. Henrion, Paige Cundiff.

**Visualization:** Ivan Carcamo-Orive, Marc Y. R. Henrion, Kuixi Zhu.

**Writing – original draft:** Ivan Carcamo-Orive, Marc Y. R. Henrion, Kuixi Zhu, Sara Moein, Joshua W. Knowles, Rui Chang.

**Writing – review & editing:** Ivan Carcamo-Orive, Marc Y. R. Henrion, Eric E. Schadt, Thomas Quertermous, Joshua W. Knowles, Rui Chang.

## References

1. Einhorn DRG, Cobin RH, Ford E, Ganda OP, Handelsman Y, Hellman R, et al. American College of Endocrinology position statement on the insulin resistance syndrome. *Endocr Pract.* 2003; 9:237–52. PMID: [12924350](#)
2. Bonnefond A, Froguel P. Rare and common genetic events in type 2 diabetes: what should biologists know? *Cell Metab.* 2015; 21(3):357–68. <https://doi.org/10.1016/j.cmet.2014.12.020> PMID: [25640731](#)
3. Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet.* 2012; 44(9):991–1005. <https://doi.org/10.1038/ng.2385> PMID: [22885924](#)
4. Dimas AS, Lagou V, Barker A, Knowles JW, Magi R, Hivert MF, et al. Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes.* 2014; 63(6):2158–71. <https://doi.org/10.2337/db13-0949> PMID: [24296717](#)
5. Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proenca C, et al. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet.* 2009; 41(10):1110–5. <https://doi.org/10.1038/ng.443> PMID: [19734900](#)
6. Knowles JW, Xie W, Zhang Z, Chennamsetty I, Assimes TL, Paananen J, et al. Identification and validation of N-acetyltransferase 2 as an insulin sensitivity gene. *J Clin Invest.* 2015; 125(4):1739–51. <https://doi.org/10.1172/JCI74692> PMID: [25798622](#)
7. Knowles JW, Assimes TL, Tsao PS, Natali A, Mari A, Quertermous T, et al. Measurement of insulin-mediated glucose uptake: direct comparison of the modified insulin suppression test and the euglycemic, hyperinsulinemic clamp. *Metabolism: clinical and experimental.* 2013; 62(4):548–53. <https://doi.org/10.1016/j.metabol.2012.10.002> PMID: [23151437](#)
8. Ferrannini E, Mari A. How to measure insulin sensitivity. *J Hypertens.* 1998; 16(7):895–906. <https://doi.org/10.1097/00004872-199816070-00001> PMID: [9794728](#)
9. Carcamo-Orive I, Hoffman GE, Cundiff P, Beckmann ND, D'Souza SL, Knowles JW, et al. Analysis of Transcriptional Variability in a Large Human iPSC Library Reveals Genetic and Non-genetic Determinants of Heterogeneity. *Cell Stem Cell.* 2017; 20(4):518–32 e9. <https://doi.org/10.1016/j.stem.2016.11.005> PMID: [28017796](#)
10. Toledo JB, Arnold M, Kastenmuller G, Chang R, Baillie RA, Han X, et al. Metabolic network failures in Alzheimer's disease: A biochemical road map. *Alzheimers Dement.* 2017; 13(9):965–84. <https://doi.org/10.1016/j.jalz.2017.01.020> PMID: [28341160](#)
11. Chang R, Karr JR, Schadt EE. Causal inference in biology networks with integrated belief propagation. *Pac Symp Biocomput.* 2015:359–70. PMID: [25592596](#)
12. Peters LA, Perrigoue J, Mortha A, Iuga A, Song WM, Neiman EM, et al. A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nat Genet.* 2017; 49(10):1437–49. <https://doi.org/10.1038/ng.3947> PMID: [28892060](#)
13. Lee S, Zhang C, Kilicarslan M, Piening BD, Bjornson E, Hallstrom BM, et al. Integrated Network Analysis Reveals an Association between Plasma Mannose Levels and Insulin Resistance. *Cell Metab.* 2016; 24(1):172–84. <https://doi.org/10.1016/j.cmet.2016.05.026> PMID: [27345421](#)
14. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9:559. <https://doi.org/10.1186/1471-2105-9-559> PMID: [19114008](#)
15. Yip J, Facchini FS, Reaven GM. Resistance to insulin-mediated glucose disposal as a predictor of cardiovascular disease. *J Clin Endocrinol Metab.* 1998; 83(8):2773–6. <https://doi.org/10.1210/jcem.83.8.5005> PMID: [9709945](#)
16. Facchini FS, Hua N, Abbasi F, Reaven GM. Insulin resistance as a predictor of age-related diseases. *J Clin Endocrinol Metab.* 2001; 86(8):3574–8. <https://doi.org/10.1210/jcem.86.8.7763> PMID: [11502781](#)
17. Yeni-Komshian H, Carantoni M, Abbasi F, Reaven GM. Relationship between several surrogate estimates of insulin resistance and quantification of insulin-mediated glucose disposal in 490 healthy nondiabetic volunteers. *Diabetes Care.* 2000; 23(2):171–5. <https://doi.org/10.2337/diacare.23.2.171> PMID: [10868826](#)

18. Iovino S, Burkart AM, Kriauciunas K, Warren L, Hughes KJ, Molla M, et al. Genetic insulin resistance is a potent regulator of gene expression and proliferation in human iPS cells. *Diabetes*. 2014; 63(12):4130–42. <https://doi.org/10.2337/db14-0109> PMID: 25059784
19. Iovino S, Burkart AM, Warren L, Patti ME, Kahn CR. Myotubes derived from human-induced pluripotent stem cells mirror in vivo insulin resistance. *Proc Natl Acad Sci U S A*. 2016; 113(7):1889–94. <https://doi.org/10.1073/pnas.1525665113> PMID: 26831110
20. Linderman M, Zhang B. coexpp: Large-scale Coexpression network creation and manipulation using WGCNA. R package version 0.1.0. <https://bitbucket.org/multiscale/coexpp>. 2011.
21. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102(43):15545–50. <https://doi.org/10.1073/pnas.0506580102> PMID: 16199517
22. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics*. 2003; 34(3):267–73. <https://doi.org/10.1038/ng1180> PMID: 12808457
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25(1):25–9. <https://doi.org/10.1038/75556> PMID: 10802651
24. Gene Ontology C. Gene Ontology Consortium: going forward. *Nucleic Acids Res*. 2015; 43(Database issue):D1049–56. <https://doi.org/10.1093/nar/gku1179> PMID: 25428369
25. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res*. 2013; 41(Database issue):D793–800. <https://doi.org/10.1093/nar/gks1055> PMID: 23143270
26. Zhang B, Zhu J. Identification of Key Causal Regulators in Gene Networks. *Proceedings of the World Congress on Engineering & Computer Science*. 2013; 2.
27. Wu Q, Ortegon AM, Tsang B, Doege H, Feingold KR, Stahl A. FATP1 is an insulin-sensitive fatty acid transporter involved in diet-induced obesity. *Mol Cell Biol*. 2006; 26(9):3455–67. <https://doi.org/10.1128/MCB.26.9.3455-3467.2006> PMID: 16611988
28. Kim JK, Gimeno RE, Higashimori T, Kim HJ, Choi H, Punreddy S, et al. Inactivation of fatty acid transport protein 1 prevents fat-induced insulin resistance in skeletal muscle. *J Clin Invest*. 2004; 113(5):756–63. <https://doi.org/10.1172/JCI18917> PMID: 14991074
29. Choi JW, Jo A, Kim M, Park HS, Chung SS, Kang S, et al. BNIP3 is essential for mitochondrial bioenergetics during adipocyte remodelling in mice. *Diabetologia*. 2016; 59(3):571–81. <https://doi.org/10.1007/s00125-015-3836-9> PMID: 26693709
30. Glick D, Zhang W, Beaton M, Marsboom G, Gruber M, Simon MC, et al. BNIP3 regulates mitochondrial function and lipid metabolism in the liver. *Mol Cell Biol*. 2012; 32(13):2570–84. <https://doi.org/10.1128/MCB.00167-12> PMID: 22547685
31. Tol MJ, Ottenhoff R, van Eijk M, Zelcer N, Aten J, Houten SM, et al. A PPAR $\gamma$ -Bnip3 Axis Couples Adipose Mitochondrial Fusion-Fission Balance to Systemic Insulin Sensitivity. *Diabetes*. 2016; 65(9):2591–605. <https://doi.org/10.2337/db16-0243> PMID: 27325287
32. Grice BA, & Elmendorf J. S. New aspects of cellular cholesterol regulation on blood glucose control—review and perspective on the impact of statin medications on metabolic health. *US Endocrinology*. 2017; 13(2):63–8.
33. Ingelsson E, Knowles JW. Leveraging Human Genetics to Understand the Relation of LDL Cholesterol with Type 2 Diabetes. *Clin Chem*. 2017; 63(7):1187–9. <https://doi.org/10.1373/clinchem.2016.268565> PMID: 28515096
34. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. *N Engl J Med*. 2016; 375(22):2144–53. <https://doi.org/10.1056/NEJMoa1604304> PMID: 27959767
35. Swerdlow DI, Preiss D, Kuchenbaecker KB, Holmes MV, Engmann JE, Shah T, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. *Lancet*. 2015; 385(9965):351–61. [https://doi.org/10.1016/S0140-6736\(14\)61183-1](https://doi.org/10.1016/S0140-6736(14)61183-1) PMID: 25262344
36. Stone NJ, Robinson JG, Lichtenstein AH, Bairey Merz CN, Blum CB, Eckel RH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014; 129(25 Suppl 2):S1–45.
37. Forst CV, Zhou B, Wang M, Chou TW, Mason G, Song WM, et al. Integrative gene network analysis identifies key signatures, intrinsic networks and host factors for influenza virus A infections. *NPJ Syst Biol Appl*. 2017; 3:35. <https://doi.org/10.1038/s41540-017-0036-x> PMID: 29214055

38. Kohli P, Waters DD, Nemr R, Arsenault BJ, Messig M, DeMicco DA, et al. Risk of new-onset diabetes and cardiovascular risk reduction from high-dose statin therapy in pre-diabetics and non-pre-diabetics: an analysis from TNT and IDEAL. *J Am Coll Cardiol*. 2015; 65(4):402–4. <https://doi.org/10.1016/j.jacc.2014.10.053> PMID: 25634843
39. Fischer-Posovszky P, Newell FS, Wabitsch M, Tornqvist HE. Human SGBS cells—a unique tool for studies of human fat cell biology. *Obes Facts*. 2008; 1(4):184–9. <https://doi.org/10.1159/000145784> PMID: 20054179
40. Rokach O, Ullrich ND, Rausch M, Mouly V, Zhou H, Muntoni F, et al. Establishment of a human skeletal muscle-derived cell line: biochemical, cellular and electrophysiological characterization. *The Biochemical journal*. 2013; 455(2):169–77. <https://doi.org/10.1042/BJ20130698> PMID: 23905709
41. Kohli P, Knowles JW, Sarraju A, Waters DD, Reaven G. Metabolic Markers to Predict Incident Diabetes Mellitus in Statin-Treated Patients (from the Treating to New Targets and the Stroke Prevention by Aggressive Reduction in Cholesterol Levels Trials). *Am J Cardiol*. 2016; 118(9):1275–81. <https://doi.org/10.1016/j.amjcard.2016.07.054> PMID: 27614854
42. Collins R, Reith C, Emberson J, Armitage J, Baigent C, Blackwell L, et al. Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet*. 2016; 388(10059):2532–61. [https://doi.org/10.1016/S0140-6736\(16\)31357-5](https://doi.org/10.1016/S0140-6736(16)31357-5) PMID: 27616593
43. Preiss D, Seshasai SR, Welsh P, Murphy SA, Ho JE, Waters DD, et al. Risk of incident diabetes with intensive-dose compared with moderate-dose statin therapy: a meta-analysis. *JAMA*. 2011; 305(24):2556–64. <https://doi.org/10.1001/jama.2011.860> PMID: 21693744
44. Sattar N, Preiss D, Murray HM, Welsh P, Buckley BM, de Craen AJ, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet*. 2010; 375(9716):735–42. [https://doi.org/10.1016/S0140-6736\(09\)61965-6](https://doi.org/10.1016/S0140-6736(09)61965-6) PMID: 20167359
45. Cederberg H, Stancakova A, Yaluri N, Modi S, Kuusisto J, Laakso M. Increased risk of diabetes with statin treatment is associated with impaired insulin sensitivity and insulin secretion: a 6 year follow-up study of the METSIM cohort. *Diabetologia*. 2015; 58(5):1109–17. <https://doi.org/10.1007/s00125-015-3528-5> PMID: 25754552
46. Ridker PM, Pradhan A, MacFadyen JG, Libby P, Glynn RJ. Cardiovascular benefits and diabetes risks of statin therapy in primary prevention: an analysis from the JUPITER trial. *Lancet*. 2012; 380(9841):565–71. [https://doi.org/10.1016/S0140-6736\(12\)61190-8](https://doi.org/10.1016/S0140-6736(12)61190-8) PMID: 22883507
47. Besseling J, Kastelein JJ, Defesche JC, Hutten BA, Hovingh GK. Association between familial hypercholesterolemia and prevalence of type 2 diabetes mellitus. *JAMA*. 2015; 313(10):1029–36. <https://doi.org/10.1001/jama.2015.1206> PMID: 25756439
48. Brault M, Ray J, Gomez YH, Mantzoros CS, Daskalopoulou SS. Statin treatment and new-onset diabetes: a review of proposed mechanisms. *Metabolism: clinical and experimental*. 2014; 63(6):735–45. <https://doi.org/10.1016/j.metabol.2014.02.014> PMID: 24641882
49. Kain V, Kapadia B, Misra P, Saxena U. Simvastatin may induce insulin resistance through a novel fatty acid mediated cholesterol independent mechanism. *Sci Rep*. 2015; 5:13823. <https://doi.org/10.1038/srep13823> PMID: 26345110
50. Elfakhani M, Torabi S, Hussein D, Mills N, Verbeck GF, Mo H. Mevalonate deprivation mediates the impact of lovastatin on the differentiation of murine 3T3-F442A preadipocytes. *Exp Biol Med (Maywood)*. 2014; 239(3):293–301. <https://doi.org/10.1177/1535370213517614> PMID: 24477821
51. Nakata M, Nagasaka S, Kusaka I, Matsuoka H, Ishibashi S, Yada T. Effects of statins on the adipocyte maturation and expression of glucose transporter 4 (SLC2A4): implications in glycaemic control. *Diabetologia*. 2006; 49(8):1881–92. <https://doi.org/10.1007/s00125-006-0269-5> PMID: 16685502
52. Parpal S, Karlsson M, Thorn H, Stralfors P. Cholesterol depletion disrupts caveolae and insulin receptor signaling for metabolic control via insulin receptor substrate-1, but not for mitogen-activated protein kinase control. *J Biol Chem*. 2001; 276(13):9670–8. <https://doi.org/10.1074/jbc.M007454200> PMID: 11121405
53. Chamberlain LH. Inhibition of isoprenoid biosynthesis causes insulin resistance in 3T3-L1 adipocytes. *FEBS Lett*. 2001; 507(3):357–61. [https://doi.org/10.1016/s0014-5793\(01\)03007-1](https://doi.org/10.1016/s0014-5793(01)03007-1) PMID: 11696371
54. Takaguri A, Satoh K, Itagaki M, Tokumitsu Y, Ichihara K. Effects of atorvastatin and pravastatin on signal transduction related to glucose uptake in 3T3L1 adipocytes. *J Pharmacol Sci*. 2008; 107(1):80–9. <https://doi.org/10.1254/jphs.fp0072403> PMID: 18469500
55. Li W, Liang X, Zeng Z, Yu K, Zhan S, Su Q, et al. Simvastatin inhibits glucose uptake activity and GLUT4 translocation through suppression of the IR/IRS-1/Akt signaling in C2C12 myotubes. *Biomed Pharmacother*. 2016; 83:194–200. <https://doi.org/10.1016/j.biopha.2016.06.029> PMID: 27470565
56. Mallinson JE, Constantin-Teodosiu D, Sidaway J, Westwood FR, Greenhaff PL. Blunted Akt/FOXO signalling and activation of genes controlling atrophy and fuel use in statin myopathy. *J Physiol*. 2009; 587(1):219–30. <https://doi.org/10.1113/jphysiol.2008.164699> PMID: 19001041

57. Larsen S, Stride N, Hey-Mogensen M, Hansen CN, Bang LE, Bundgaard H, et al. Simvastatin effects on skeletal muscle: relation to decreased mitochondrial function and glucose intolerance. *J Am Coll Cardiol.* 2013; 61(1):44–53. <https://doi.org/10.1016/j.jacc.2012.09.036> PMID: 23287371
58. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
59. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014; 30(7):923–30. <https://doi.org/10.1093/bioinformatics/btt656> PMID: 24227677
60. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2015.
61. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616> PMID: 19910308
62. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015; 43(7):e47. <https://doi.org/10.1093/nar/gkv007> PMID: 25605792
63. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010; 11(3):R25. <https://doi.org/10.1186/gb-2010-11-3-r25> PMID: 20196867
64. Mostafavi S, Battle A, Zhu X, Urban AE, Levinson D, Montgomery SB, et al. Normalizing RNA-sequencing data by modeling hidden covariates with prior knowledge. *PLoS One.* 2013; 8(7):e68141. <https://doi.org/10.1371/journal.pone.0068141> PMID: 23874524
65. Choi SH, Labadorf AT, Myers RH, Lunetta KL, Dupuis J, DeStefano AL. Evaluation of logistic regression models and effect of covariates for case-control study in RNA-Seq analysis. *BMC Bioinformatics.* 2017; 18(1):91. <https://doi.org/10.1186/s12859-017-1498-y> PMID: 28166718
66. Hoffman GE, Schadt EE. variancePartition: Quantifying and interpreting drivers of variation in complex gene expression studie. *bioRxiv.* 2016.
67. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2012; 9(2):179–81.
68. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics.* 2009; 5(6):e1000529. <https://doi.org/10.1371/journal.pgen.1000529> PMID: 19543373
69. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015; 526(7571):68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245
70. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics.* 2012; 28(10):1353–8. <https://doi.org/10.1093/bioinformatics/bts163> PMID: 22492648
71. Stegle O, Parts L, Durbin R, Winn J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLoS Comput Biol.* 2010; 6(5):e1000770. <https://doi.org/10.1371/journal.pcbi.1000770> PMID: 20463871
72. Needham CJ, Bradford JR, Bulpitt AJ, Westhead DR. Inference in Bayesian networks. *Nat Biotechnol.* 2006; 24(1):51–3. <https://doi.org/10.1038/nbt0106-51> PMID: 16404397
73. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 2008; 6(5):e107. <https://doi.org/10.1371/journal.pbio.0060107> PMID: 18462017
74. Sivendran S, Chang R, Pham L, Phelps RG, Harcharik ST, Hall LD, et al. Dissection of immune gene networks in primary melanoma tumors critical for antitumor surveillance of patients with stage II-III resectable disease. *J Invest Dermatol.* 2014; 134(8):2202–11. <https://doi.org/10.1038/jid.2014.85> PMID: 24522433
75. Liu F, Zhang SW, Guo WF, Wei ZG, Chen L. Inference of Gene Regulatory Network Based on Local Bayesian Networks. *PLoS Comput Biol.* 2016; 12(8):e1005024. <https://doi.org/10.1371/journal.pcbi.1005024> PMID: 27479082
76. Petyuk VA, Chang R, Ramirez-Restrepo M, Beckmann ND, Henrion MYR, Piehowski PD, et al. The human brainome: network analysis identifies HSPA2 as a novel Alzheimer's disease target. *Brain.* 2018; 141(9):2721–39. <https://doi.org/10.1093/brain/awy215> PMID: 30137212
77. Chang R, Shoemaker R, Wang W. A novel knowledge-driven systems biology approach for phenotype prediction upon genetic intervention. *IEEE/ACM Trans Comput Biol Bioinform.* 2011; 8(5):1170–82. <https://doi.org/10.1109/TCBB.2011.18> PMID: 21282866



78. Chang R, Shoemaker R, Wang W. Systematic search for recipes to generate induced pluripotent stem cells. *PLoS Comput Biol*. 2011; 7(12):e1002300. <https://doi.org/10.1371/journal.pcbi.1002300> PMID: 22215993
79. Pryciak PM. Systems biology. Customized signaling circuits. *Science*. 2008; 319(5869):1489–90. <https://doi.org/10.1126/science.1156414> PMID: 18339925
80. Myers AJ, Chang R, Petyuk VA, Ramirez-Restrepo M, Beckmann ND, Henrion MYR, et al. THE HUMAN BRAINOME: HUMAN BRAIN GENOME, TRANSCRIPTOME, AND PROTEOME INTEGRATION. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* 2016. p. 237–8.
81. Petyuk VA, Chang R, Ramirez-Restrepo M, Beckmann ND, Henrion MYR, Piehowski PD, et al. The human brainome: network analysis identifies HSPA2 as a novel Alzheimer's disease target. *Brain*. 2018; 141(9):2721–39. <https://doi.org/10.1093/brain/awy215> PMID: 30137212
82. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet*. 2005; 37(7):710–7. <https://doi.org/10.1038/ng1589> PMID: 15965475
83. Lorsch ZS, Hamilton PJ, Ramakrishnan A, Parise EM, Sallery M, Wright WJ, et al. Stress resilience is promoted by a Zfp189-driven transcriptional network in prefrontal cortex. *Nat Neurosci*. 2019; 22(9):1413–23. <https://doi.org/10.1038/s41593-019-0462-8> PMID: 31427770
84. Beckmann ND, Lin W-J, Wang M, Cohain AT, Charney AW, Wang P, et al. Multiscale causal network models of Alzheimer's disease identify VGF as a key regulator of disease. *Nat Commun*. 2020; 11:3942 <https://doi.org/10.1038/s41467-020-17405-z> PMID: 32770063
85. Koh HJ, Lee SM, Son BG, Lee SH, Ryoo ZY, Chang KT, et al. Cytosolic NADP<sup>+</sup>-dependent isocitrate dehydrogenase plays a key role in lipid metabolism. *J Biol Chem*. 2004; 279(38):39968–74. <https://doi.org/10.1074/jbc.M402260200> PMID: 15254034
86. Nam WS, Park KM, Park JW. RNA interference targeting cytosolic NADP<sup>+</sup>-dependent isocitrate dehydrogenase exerts anti-obesity effect in vitro and in vivo. *Biochim Biophys Acta*. 2012; 1822(8):1181–8. <https://doi.org/10.1016/j.bbadis.2012.04.003> PMID: 22542506
87. Shin CY, Jung MY, Lee IK, Son M, Kim DS, Lim JI, et al. Anti-diabetic effects of DA-11004, a synthetic IDPc inhibitor in high fat high sucrose diet-fed C57BL/6J mice. *Arch Pharm Res*. 2004; 27(1):48–52. <https://doi.org/10.1007/BF02980045> PMID: 14969338
88. Chu B, Wu T, Miao L, Mei Y, Wu M. MiR-181a regulates lipid metabolism via IDH1. *Sci Rep*. 2015; 5:8801. <https://doi.org/10.1038/srep08801> PMID: 25739786