Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition

Kerrin S. Small1,20*, Marijana Todorčević22,20, Mete Civelek3,4,20, Julia S. El-Sayed Moustafa1,20, Xiao Wang5,20, Michelle M. Simon6, Juan Fernandez-Tajes7, Anubha Mahajan8,9, Momoko Horikoshi2,7, Alison Hugill8,11, Gudmar Thorleifsson12, Avanthi Raghavan13, Qiu wrong Ding14, Andrew P. Morris215, Jordana T. Bell1, Unnur Thorsteinsdottir12,16, Kari Stefansson12,16, Markku Laakso17, Ingrid Dahlman18, Peter Arner18, Anna L. Gloyn19,21, Aldons J. Lusis3,19,12, Roger D. Cox8,21, Fredrik Karpe2,9,21 and Mark I. McCarthy2,7,9,21*

Individual risk of type 2 diabetes (T2D) is modified by perturbations to the mass, distribution and function of adipose tissue. To investigate the mechanisms underlying these associations, we explored the molecular, cellular and whole-body effects of T2D-associated alleles near KLF14. We show that KLF14 diabetes-risk alleles act in adipose tissue to reduce KLF14 expression and modulate, in trans, the expression of 385 genes. We demonstrate, in human cellular studies, that reduced KLF14 expression increases pre-adipocyte proliferation but disrupts lipogenesis, and in mice, that adipose tissue-specific deletion of KLF14 partially recapitulates the human phenotype of insulin resistance, dyslipidemia and T2D. We show that carriers of the KLF14 T2D risk allele shift body fat from gynoid stores to abdominal stores and display a marked increase in adipocyte cell size, and that these effects on fat distribution, and the T2D association, are female specific. The metabolic risk associated with variation at this imprinted locus depends on the sex both of the subject and of the parent from whom the risk allele derives.

The replicated genome-wide significant T2D association signal at chr7q32.3 maps to a 45-kb recombination interval that extends from 3 kb to 48 kb upstream of KLF14 (Fig. 1a–c). In previous work based on microarray–derived RNA-expression data, KLF14, which encodes an imprinted transcript, was exposed as the likely cis-effector gene for this locus in subcutaneous adipose tissue and was shown to be a trans regulator of a program of adipose tissue expression. The KLF family of zinc-finger binding proteins has wide-ranging regulatory roles in biological processes such as proliferation, differentiation and growth. However, little is known about KLF14, a single-exon gene whose transcription is limited to the maternally inherited chromosome in embryonic, extra-embryonic and adult tissue in humans and mice.

Results
Adipose-specific regulation of KLF14 mediates the T2D association. Using RNA-based next-generation sequencing (RNA-seq) data from subcutaneous adipose biopsies collected from 766 female twin members of the TwinsUK cohort, we confirmed the cis-expression quantitative trait locus (eQTL) (using the SNP rs4731702 T2D risk allele C as the reference; \( \beta = -0.56; P = 1.8 \times 10^{-36} \)) and identified a 38-fold-expanded trans network of 385 genes (false-discovery rate (FDR) <5%) (Fig. 2 and Supplementary Table 1). The cis and trans effects at KLF14 were robust and were replicated in three independent studies of expression in subcutaneous adipose tissue (MGH8, METSIM9 and deCODE12,16) (Fig. 2c and Supplementary Tables 1 and 2). Despite detectable KLF14 expression in a range of tissues, these cis and trans associations were completely adipose specific, with

---

1Department of Twin Research and Genetic Epidemiology, King’s College London, London, UK. 2Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Oxford, UK. 3Center for Public Health Genomics, Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, USA. 4Department of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. 5Cardiovascular Institute, Department of Medicine, Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. 6Biocomputing, Medical Research Council Harwell Institute, Oxford, UK. 7Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. 8Genetics of Type 2 Diabetes, Medical Research Council Harwell Institute, Oxford, UK. 9Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK. 10Department of Human Genetics, Medicine, Department of Genetics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. 11Department of Biostatistics, University of Liverpool, Liverpool, UK. 12Department of Twin Research and Genetic Epidemiology, King’s College London, London, UK. 13Department of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. 14Faculty of Medicine, University of Iceland, Reykjavik, Iceland. 15Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. 16Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden. 17Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA, USA. 18These authors contributed equally: Kerrin S. Small, Marijana Todorčević, Mete Civelek, Julia S. El-Sayed Moustafa and Xiao Wang. 19These authors jointly directed this work: Kiran Musunuru, Aldons J. Lusis, Roger D. Cox, Fredrik Karpe and Mark I. McCarthy. *e-mail: kerrin.small@kcl.ac.uk; mark.mcgarthy@drl.ox.ac.uk

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.
no eQTL signal evident in the following: (i) skin, whole blood or lymphoblastoid cell lines from the same TwinsUK subjects (Figs. 1f and 2d); (ii) T2D-relevant tissues, such as muscle, liver or islet; or (iii) the broader coverage represented in eQTL datasets, such as GTEx (Supplementary Table 3). *KLF14* is imprinted and, as with its T2D association, the *KLF14* cis and trans eQTLs were maternal
specific—the paternally inherited allele had no effect on KLF14 expression (Supplementary Fig. 1 and Supplementary Table 4). All other genes within 300 kb of the interval are also imprinted but paternally expressed, confirming KLF14 as the likely mediator of the T2D association at this locus.

We first sought to refine the location of the causal variant(s) responsible for the associations noted above. The T2D association could not be resolved beyond a set of 29 SNPs in high mutual linkage disequilibrium (LD) ($r^2 > 0.94$ in the UK10K project); the apparent European specificity of the T2D association [15] precludes trans-ethnic fine mapping. The ancestral T2D risk allele varies in global frequency within the 1000 Genomes Project populations (EUR, 54%; ASN, 69%; AMR, 58%; AFR, 73%), but we detected no evidence of positive selection (Supplementary Fig. 2 and Supplementary Note 1). We used chromatin-state maps to annotate the associated interval and identified a ~1.6-kb enhancer ~5 kb upstream of KLF14. This enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity: in ChromHMM software predictions [16], the enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity. The enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity. The enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity. The enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity. The enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity.

We used chromatin-state maps to annotate the associated interval and identified a ~1.6-kb enhancer ~5 kb upstream of KLF14. This enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity: in ChromHMM software predictions [16] and Roadmap [17], and it is annotated as active in 10 (out of 127) cell types and tissues, 3 of which are derived from adipose tissue, including adipocytes and adipose tissue-derived mesenchymal stem cells (Fig. 1d). The enhancer is also active in certain adipose active enhancers in the region.

The enhancer is also active in certain adipose active enhancers in the region. We used chromatin-state maps to annotate the associated interval and identified a ~1.6-kb enhancer ~5 kb upstream of KLF14. This enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity: in ChromHMM software predictions [16] and Roadmap [17], and it is annotated as active in 10 (out of 127) cell types and tissues, 3 of which are derived from adipose tissue, including adipocytes and adipose tissue-derived mesenchymal stem cells (Fig. 1d). The enhancer is also active in certain adipose active enhancers in the region.

The enhancer is also active in certain adipose active enhancers in the region. We used chromatin-state maps to annotate the associated interval and identified a ~1.6-kb enhancer ~5 kb upstream of KLF14. This enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity: in ChromHMM software predictions [16] and Roadmap [17], and it is annotated as active in 10 (out of 127) cell types and tissues, 3 of which are derived from adipose tissue, including adipocytes and adipose tissue-derived mesenchymal stem cells (Fig. 1d). The enhancer is also active in certain adipose active enhancers in the region.

The enhancer is also active in certain adipose active enhancers in the region. We used chromatin-state maps to annotate the associated interval and identified a ~1.6-kb enhancer ~5 kb upstream of KLF14. This enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity: in ChromHMM software predictions [16] and Roadmap [17], and it is annotated as active in 10 (out of 127) cell types and tissues, 3 of which are derived from adipose tissue, including adipocytes and adipose tissue-derived mesenchymal stem cells (Fig. 1d). The enhancer is also active in certain adipose active enhancers in the region.

The enhancer is also active in certain adipose active enhancers in the region. We used chromatin-state maps to annotate the associated interval and identified a ~1.6-kb enhancer ~5 kb upstream of KLF14. This enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity: in ChromHMM software predictions [16] and Roadmap [17], and it is annotated as active in 10 (out of 127) cell types and tissues, 3 of which are derived from adipose tissue, including adipocytes and adipose tissue-derived mesenchymal stem cells (Fig. 1d). The enhancer is also active in certain adipose active enhancers in the region.

The enhancer is also active in certain adipose active enhancers in the region. We used chromatin-state maps to annotate the associated interval and identified a ~1.6-kb enhancer ~5 kb upstream of KLF14. This enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity: in ChromHMM software predictions [16] and Roadmap [17], and it is annotated as active in 10 (out of 127) cell types and tissues, 3 of which are derived from adipose tissue, including adipocytes and adipose tissue-derived mesenchymal stem cells (Fig. 1d). The enhancer is also active in certain adipose active enhancers in the region.

The enhancer is also active in certain adipose active enhancers in the region. We used chromatin-state maps to annotate the associated interval and identified a ~1.6-kb enhancer ~5 kb upstream of KLF14. This enhancer encompasses 5 of the 29 associated variants and shows marked tissue specificity: in ChromHMM software predictions [16] and Roadmap [17], and it is annotated as active in 10 (out of 127) cell types and tissues, 3 of which are derived from adipose tissue, including adipocytes and adipose tissue-derived mesenchymal stem cells (Fig. 1d). The enhancer is also active in certain adipose active enhancers in the region.
As methylation is one of the key processes related to enhancer function, we explored population-level methylation data at this locus in many of the same TwinsUK subjects. The T2D risk haplotype was associated with increased methylation levels at Illumina 450K array probe cg02385110, ~3 kb upstream of KLF14, in subcutaneous adipose tissue (\(n=603\), \(P=2.2\times 10^{-7}\), and \(\beta=0.01\) for rs4731702) but not whole blood (\(n=309\); \(P=0.69\)) or skin (\(n=437\); \(P=0.39\)) (Fig. 1c–i). There was a consistent direction of effect across these analyses in adipose: the T2D-risk haplotype was associated with increased methylation and decreased RNA expression, and cg02385110 methylation and KLF14 expression were negatively correlated. A second probe (cg08097417) located at the KLF14 transcription start site has been associated with age in whole blood\(^a\) and adipose-tissue samples\(^b\). However, this probe lies outside the T2D-association interval, and cg08097417 methylation was not related to KLF14 expression (\(P=0.36\) or risk haplotype (\(P=0.99\), indicating that age-related variability of cg08097417 is unrelated to KLF14 expression or disease pathogenesis (Supplementary Fig. 3). Instead, we conclude that T2D-associated risk attributable to this locus is probably a consequence of sequence variation at the adipose enhancer upstream of KLF14 and is marked by altered methylation.

The KLF14 variants regulate a large adipose tissue–specific trans network. The trans network regulated by the KLF14 variants is remarkable both for its size and its robust replication. Consistent with the known function of KLF14 as both a transcriptional activator and a transcriptional repressor, the trans associations included both positive effects and negative effects (Fig. 2a and Supplementary Table 1). Mediation analysis applied to the RNA-expression data supported the proposal of a causal role for KLF14 expression in the regulation of many of the trans genes (161 of 385 genes passed Bonferroni-corrected Sobel’s mediation: \(P<1.3\times 10^{-3}\)).

The principal trans-regulatory mechanism appeared to be direct interaction of KLF14 with trans-gene cis-regulatory elements; the 20-kb regions upstream of the 385 trans genes showed enrichment both for KLF14 binding peaks in empirical data obtained by chromatin immunoprecipitation followed by deep sequencing (ChIP-seq)\(^c\) of HEK293T human embryonic kidney cells (\(P=2\times 10^{-4}\)) and the presence of the proposed KLF14 binding motif (normalized enrichment score (NES) = 4.17; \(P=1.5\times 10^{-3}\)). We assessed functional annotation of the trans genes using TopGene\(^d\) and found the subset of 177 trans genes with KLF14-binding motifs to be enriched for the categories ‘metabolic pathways’ (\(P=1.2\times 10^{-3}\) (Bonferroni corrected)) and ‘binding by PPARG and RXRA during adipocyte differentiation’ (\(P=3.9\times 10^{-4}\) (Bonferroni corrected)) (Supplementary Table 5). Further sub-setting defining 122 trans genes enriched for both KLF14-binding motifs and KLF14 ChIP-seq peaks showed broadly similar functional enrichment (‘binding by PPARG and RXRA during adipocyte differentiation’, \(P=0.04\) (Bonferroni corrected); and ‘metabolic pathways’, \(P=2.6\times 10^{-4}\) (Bonferroni corrected)).

A sub-network of 18 trans genes showed enrichment for binding motifs for the transcription factor SREBF1 (\(P=4.22\times 10^{-4}\); \(P=1.2\times 10^{-3}\)), 11 of which did not have KLF14 motifs (Fig. 2b). The sub-network of SREBF1 motif–containing trans genes showed enrichment for the category ‘cholesterol biosynthesis and lipid metabolism’ (\(P=2.8\times 10^{-7}\) (Bonferroni correction)) (Supplementary Table 5). SREBF1, which encodes a transcription factor involved in cholesterol homeostasis, is itself a trans gene directly regulated by KLF14 and appears to act as an intermediary in the regulation of this sub-network.

The human trans network includes several genes encoding molecules with functions that make them attractive candidates for mediating the range of KLF14-associated phenotypes. These include SLC2A4 and IDE, which encode the glucose transporter GLUT4 and insulin-degrading enzyme IDE, respectively. GLUT4 mediates glucose uptake in adipose tissue and skeletal muscle, among other tissues, and an adipose tissue–specific reduction in the expression of SLC2A4 has been reported in T2D\(^c\). IDE degrades peptides, including insulin, glucagon and amylin, and maps to a long-established T2D GWAS interval\(^c\). Notably, at multiple trans genes, we identified adipose-tissue cis-eQTLs coincident with metabolic-trait GWAS SNPs (\(r^2>0.8\), including three known T2D-associated loci (STARD10, C6orf57 and CDK2AP1) (Supplementary Table 6). This shows that local regulation of these genes can mediate susceptibility to T2D independently of trans regulation via KLF14 and indicates that the phenotypic consequences of KLF14 variation are probably mediated by multiple genes within the trans network.

### Table 1 | SNP rs4731702 is associated with insulin resistance, lipid and body-composition phenotypes

| GWAS citation | Type 2 diabetes | Fasting insulin | Fasting glucose | HDL | TG | LDL | BMI | WHR | Waist | Hip |
|---------------|----------------|----------------|----------------|-----|----|-----|-----|-----|-------|-----|
| Morris, 2012\(^a\) | 1.10 (1.06–1.15) | 0.019 | 0.007 | 0.155 | 0.16 | 0.09 | 0.009 | 0.006 | 0.009 | 0.017 |
| Horikoshi, 2015\(^b\) | 1.10 (1.06–1.15) | 0.019 | 0.007 | 0.155 | 0.16 | 0.09 | 0.009 | 0.006 | 0.009 | 0.017 |
| Horikoshi, 2015\(^b\) | 1.10 (1.06–1.15) | 0.019 | 0.007 | 0.155 | 0.16 | 0.09 | 0.009 | 0.006 | 0.009 | 0.017 |
| Teslovich, 2010\(^c\) | 1.14 (1.08–1.20) | 0.031 | 0.011 | 0.042 | 0.036 | 0.018 | 0.010 | 0.021 | 0.008 | 0.033 |
| Teslovich, 2010\(^c\) | 1.14 (1.08–1.20) | 0.031 | 0.011 | 0.042 | 0.036 | 0.018 | 0.010 | 0.021 | 0.008 | 0.033 |
| Teslovich, 2010\(^c\) | 1.14 (1.08–1.20) | 0.031 | 0.011 | 0.042 | 0.036 | 0.018 | 0.010 | 0.021 | 0.008 | 0.033 |
| Teslovich, 2010\(^c\) | 1.14 (1.08–1.20) | 0.031 | 0.011 | 0.042 | 0.036 | 0.018 | 0.010 | 0.021 | 0.008 | 0.033 |
| Teslovich, 2010\(^c\) | 1.14 (1.08–1.20) | 0.031 | 0.011 | 0.042 | 0.036 | 0.018 | 0.010 | 0.021 | 0.008 | 0.033 |
| Teslovich, 2010\(^c\) | 1.14 (1.08–1.20) | 0.031 | 0.011 | 0.042 | 0.036 | 0.018 | 0.010 | 0.021 | 0.008 | 0.033 |
| Teslovich, 2010\(^c\) | 1.14 (1.08–1.20) | 0.031 | 0.011 | 0.042 | 0.036 | 0.018 | 0.010 | 0.021 | 0.008 | 0.033 |
| Teslovich, 2010\(^c\) | 1.14 (1.08–1.20) | 0.031 | 0.011 | 0.042 | 0.036 | 0.018 | 0.010 | 0.021 | 0.008 | 0.033 |

Significance (\(p\)-value and \(P\)-value) of the association between KLF14 and various parameters (top row) in various GWASs (second row), for females and males combined (top group) and for females (middle group) and males (bottom group) separately. Triglyceride (TG) results are log-transformed values; results from ref. \(^{30}\) are taken from the stage 1 discovery, as stage 2 included samples from non-European subjects; effect sizes for HDL and LDL are reported in mmol/l; effect sizes for T2D are ORs, not \(p\)-values.
Sex-specific metabolic-trait associations implicate insulin action. Previous studies have demonstrated that T2D risk alleles at KLF14 are, in non-diabetic individuals, associated with increased fasting insulin and decreased high-density lipoprotein cholesterol (HDL-C)\(^{37,38}\), emphasizing a primary impact on insulin action rather than on insulin secretion. By collecting data from the largest available GWAS meta-analyses for multiple traits\(^{26,28–31}\), we extended the spectrum of KLF14 association to encompass a broad range of insulin-resistant, ‘metabolic-syndrome’ phenotypes, including such traits as low-density lipoprotein (LDL) cholesterol, triglycerides, waist/hip ratio (WHR) and fasting insulin\(^{32}\) (Table 1). Crucially, as demonstrated before\(^1\), the impact of KLF14 variation was far greater on fat distribution than on overall adiposity (as measured by body-mass index (BMI)), with a particularly strong association between the T2D risk allele and reduced hip circumference \((P=0.43)\). In a combination of samples from the UK Biobank, Wellcome Trust Case Control Consortium and The Resource for Genetic Epidemiology on Adult Health and Aging (GERA) (15,728 cases; 129,911 controls), rs4731702 and sex interacted to influence T2D risk \((P=6.6 \times 10^{-3})\).

The analyses noted above demonstrate that the metabolic and anthropometric consequences of KLF14 variation are dependent on both the parental origin of the risk allele and the sex of the recipient of that allele. By combining these sex-specific ORs with the known imprinting-related maternal specificity of these associations\(^1\), we estimate that the per-allele OR for T2D is \(=1.28\) in women who inherited the risk allele from their mother. In contrast, although KLF14 expression in adipose tissue was higher in females than males across all genotype classes, we found no equivalent sex difference in the magnitude of the adipose-tissue cis-eQTL effect (Fig. 3a and Supplementary Table 4). To further explore drivers of the sexual dimorphism in KLF14 expression, we compared KLF14 expression in 86 pre-menopausal females with that in 487 post-menopausal females from TwinsUK. Subjects on hormone-replacement therapy were excluded from this analysis. There was no association between KLF14 expression and menopausal status \((P=0.81)\), and the cis-eQTL effect sizes in each group were similar, which suggested that gonadal steroids do not underlie the observed sexual dimorphism in KLF14 expression levels. We conclude that the sex dependency of the whole-body phenotype is not the direct consequence of sexual dimorphism at the cis-regulatory level.

The genetic associations were consistent with our observations using gene-expression levels in the TwinsUK cohort. KLF14 expression in adipose tissue was associated with a combined insulin-resistance phenotype characterized by increased fasting insulin and triglycerides and decreased HDL-C \((P=1.08 \times 10^{-7})\). KLF14 transgenes also showed enrichment for association between trans-gene expression levels and the same combined insulin-resistance phenotype \((P=1.82 \times 10^{-7})\) (Supplementary Tables 7 and 8). We used dual-energy X-ray absorptiometry (DXA)-derived body-composition data from the TwinsUK subjects for whom gene-expression data were available to further delineate the impact of KLF14 expression on fat distribution. While there was no association between KLF14 expression in subcutaneous adipose tissue and total fat volume \((P=0.56)\) or BMI \((P=0.27)\), there were differences in the distribution of that fat, including an inverse association with abdominal visceral fat \((P=0.02)\) and with the ratio of android to gynoid fat \((P=1.7 \times 10^{-5})\). In many other insulin-resistant settings, abnormalities in fat distribution are associated with ectopic deposition of fat, including deposition in the liver, which leads to non-alcoholic fatty liver disease. However, we found no evidence that KLF14 risk variants influence fat deposition in the liver, whether measured directly \((P=0.26; n=7,176)\) or indirectly via disturbed liver function (alanine aminotransferase, \(P=0.79\); gamma-glutamyl transferase, \(P=0.89; n=61,089\))\(^1\).

These data indicate that a genetically determined reduction in the cis expression of KLF14 in adipose results in an insulin-resistant and T2D-predisposing phenotype that, in women only, is characterized by (and potentially mediated through) a shift in fat distribution from relatively inert gynoid stores to deposition in more metabolically active abdominal tissues.

**Conserved mouse phenotypes.** We generated mice carrying adipose-tissue–specific deletion of a Klf14 allele through the combination of a CRISPR-Cas9-generated allele with Klf14-flanking loxp sites and adipose-tissue–specific expression of Cre recombinase (via Adipoq) and compared those mice with their wild-type Cre-expressing colony mates. Mice of both sexes with adipose tissue–specific knockout of Klf14 (null on both maternal and paternal chromosomes) displayed a broad pattern of insulin-resistance phenotypes: HDL-C was reduced (16 weeks, significant only in
RNA-seq analysis of subcutaneous fat taken, both from mice with adipose tissue–specific knockout of Klf14 and Klf14<sup>tm1(KOMP)Vlcg</sup> mice with global deletion of Klf14, replicated some but not all of the features of the trans network seen in humans. In the mice with adipose tissue–specific knockout of Klf14 (n = 8), 87 genes were expressed differentially at an FDR q value of <0.05 (or 1,286 genes with a P value of <0.05). The latter set of (1,286 nominally significant mouse genes) showed significant enrichment for overlap with the human trans network (P < 2 × 10<sup>−6</sup>). In the Klf14<sup>tm1(KOMP)Vlcg</sup> mice with global deletion of Klf14 (n = 16), we confirmed this overlap (P < 1 × 10<sup>−4</sup> with the human trans network) in the 285 genes expressed differentially relative to their expression in control mice (FDR q < 0.05): 5′ noncoding sequences for these genes also showed enrichment for the presence of the Klf14 binding motif (P < 0.001). Although Srebfl2 (the mouse ortholog of SREBF1) is directly regulated by Klf14, we did not detect enrichment among the mouse differentially expressed genes for the Srebfl2 motif. The mouse differentially expressed genes showed enrichment for relevant functional categories, including ‘response to lipid’ (q = 4.4 × 10<sup>−4</sup>), ‘triglyceride metabolic process’ (q = 2.3 × 10<sup>−4</sup>), ‘response to steroid hormone’ (q = 2.7 × 10<sup>−4</sup>) and ‘regulation of cell proliferation’ (q = 3.2 × 10<sup>−4</sup>) (Supplementary Table 9).

Thus, adipose tissue–specific deletion of Klf14 in mice recapitulates many aspects of the phenotype associated with variants that influence adipose tissue–specific cis regulation of KLF14 in humans, but the absence of sex specificity and differences in the molecular consequences of Klf14 perturbation place limits on the relevance of mouse models for this gene.

**KLF14 impacts glucose uptake, lipogenesis and cell size.** Members of the KLF family are involved in the transcriptional control of adipocyte development and function<sup>7</sup>, but the role of KLF14 in adipogenesis is unknown. To evaluate the role of KLF14 in adipocyte development, we measured KLF14 expression in primary pre-adipocytes isolated from biopsies of human abdominal subcutaneous adipose tissue (female, n = 4; male, n = 4) during proliferation and subsequent differentiation over 14 d. Adipocyte expression of KLF14 was higher in females than in males at all time points (Fig. 3b), demonstrating that the sex-differential expression of KLF14 observed in adult biopsies is present throughout adipogenesis.

We next investigated the link between reduced expression of KLF14 and adipocyte development. Fresh adipose-tissue explants from females homozygous for the risk allele showed a 44% reduction in lipogenesis (P = 0.001; n = 132) relative to that of tissue from females homozygous for the non-risk allele, as measured by incorporation of the <sup>14</sup>C-glucose label into the triglyceride backbone (Fig. 5e). There was no difference in males (P = 0.49; n = 32) (Fig. 5e). We confirmed those findings in vitro through short hairpin RNA (shRNA)-mediated knockdown of KLF14 (40% at day 1, and 33% at day 14) (Fig. 5a) in primary pre-adipocytes isolated from abdominal subcutaneous biopsies (n = 7 females); this resulted in a 60% reduction in the accumulation of triglycerides after 14 d of differentiation (P = 0.03) (Fig. 5c). The adipocyte differentiation wholly relied on de novo lipogenesis, as the cell-culture medium contained glucose but not fatty acids. Knowing that SLC2A4 (which encodes GLUT4) is a KLF14 trans gene, we first assessed expression of GLUT4 in the KLF14<sup>shRNA</sup> knockdown adipocytes that showed lower expression as compared to control non-targeting shRNA cells (by 64% ± 7% (mean ± s.e.m.); P = 0.02) (Fig. 5a). To investigate the functional consequences of the reduction in GLUT4 expression, we quantified the insulin-stimulated uptake of glucose and found a 50% reduction in KLF14<sup>shRNA</sup> knockdown adipocytes (Fig. 5d). Concomitant with the lower glucose uptake and lipogenesis, the expression of genes encoding adipocyte-maturation factors (PLIN1 (−42% ± 8%; P = 0.02), LEP (−54% ± 13%; P = 0.03) and FITM2 (−12% ± 4%; P = 0.04)) was significantly lower in the
cells in which KLF14 was knocked down, while the levels of genes encoding classical early-differentiation, pro-adipogenic transcription factors (CEBPA and PPARG) were not changed (Fig. 5a). Knockdown of KLF14 also increased cellular proliferation, observed as reduced doubling time ($P=0.02$) (Fig. 5b). We infer that reduced levels of KLF14 result in a defect in glucose uptake, which results in impaired adipocyte lipogenesis and maturation, and we propose that the increase in proliferation results from fruitless cycling at earlier stages of differentiation.

Adipocyte size is an important marker of adipose-tissue dysfunction and metabolic disease and, in some studies, has been predictive of insulin resistance and T2D independently of obesity. We assessed the distribution of adipocyte size and cell number by histological analysis of biopsies of subcutaneous abdominal and gluteal adipose tissue from sex-, age- and BMI-matched pairs of subjects from the Oxford Biobank who were homozygous for either the risk haplotype or non-risk haplotype ($n=18$ males and 18 females). At both tissue sites, there was a marked shift toward increased adipocyte size in subjects homozygous for the T2D risk haplotype: this was observed only for women (comparison of median values: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e and Supplementary Fig. 6a–c). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e). We estimated, assuming spherical cell morphology, that for females, there was a twofold difference between the genotype groups in their adipocyte volume. Gluteal and abdominal tissue from females homozygous for the T2D risk haplotype also contained fewer total adipocytes (comparison of median values for adipocyte counts: abdominal tissue, $P=0.008$; gluteal tissue, $P=0.02$) (Fig. 6b,d,e).
which has been widely observed in epidemiological studies and accumulation associated with adipose deposition in gynoid sites, protection against the metabolic consequences of adipose-tissue transition over gluteal deposition. This results in loss of the relative tissue between subcutaneous fat depots, favoring abdominal deposition. The risk allele probably drives a shift in the distribution of adipose tissue characterized by the presence of fewer but larger adipocytes. We infer that carriers of the risk allele are obliged to favor expansion of existing adipocytes to meet lipid-storage needs, which results in suboptimal fat storage and an increased risk of insulin resistance and T2D. This hypothesis of impaired fat storage is consistent with the elevated expression of in adult adipose tissue are also implicated in a defect of adipogenesis that probably reflects impaired glucose uptake. The consequence of this defect appears to be a profound effect on the structure of subcutaneous adipose tissue characterized by the presence of fewer but larger adipocytes. We infer that carriers of the risk allele probably drives a shift in the distribution of adipose tissue between subcutaneous fat depots, favoring abdominal deposition over gluteal deposition. This results in loss of the relative protection against the metabolic consequences of adipose-tissue accumulation associated with adipose deposition in gynoid sites, which has been widely observed in epidemiological studies and is consistent with the observed association of the and human fat distribution.

**Discussion**

Our results show that T2D risk variants that lead to lower expression of in adult adipose tissue are also implicated in a defect of adipogenesis that probably reflects impaired glucose uptake. The consequence of this defect appears to be a profound effect on the structure of subcutaneous adipose tissue characterized by the presence of fewer but larger adipocytes. We infer that carriers of the risk allele are obliged to favor expansion of existing adipocytes to meet lipid-storage needs, which results in suboptimal fat storage and an increased risk of insulin resistance and T2D. This hypothesis of impaired fat storage is consistent with the elevated levels of circulating triglycerides observed in carriers of the risk allele probably drives a shift in the distribution of adipose tissue between subcutaneous fat depots, favoring abdominal deposition over gluteal deposition. This results in loss of the relative protection against the metabolic consequences of adipose-tissue accumulation associated with adipose deposition in gynoid sites, which has been widely observed in epidemiological studies and is consistent with the observed association of the and human fat distribution.

We note that the sex specificity of the risk alleles varies across biological domains: effects on adipocyte development and whole-body phenotypes were more marked in females despite the fact that cis- and trans-regulatory effects are shared between the sexes. The simplest explanation for this discordance is a threshold effect driven by the higher baseline expression of in females, although it is also possible that males lack downstream processes (for example, those related to the hormonal environment) that mediate the response to . While the --T2D association was discovered through typical GWAS approaches, the remarkable specificity of this association—with respect to sex, parent of origin, ethnicity and tissue of action—is a reminder that risk prediction, or targeted medical treatment, based on genotype alone may fail to capture highly relevant aspects of biological complexity.

**Methods**

Methods, including statements of data availability and any associated accession codes and references, are available at .

Received: 22 May 2017; Accepted: 15 February 2018; Published online: 9 April 2018

**References**

1. Kong, A. et al. Parental origin of sequence variants associated with complex diseases. Nature 462, 868–874 (2009).
2. Voight, B. F. et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet. 42, 579–589 (2010).
3. Small, K. S. et al. Identification of an imprinted master trans regulator at the locus related to multiple metabolic phenotypes. Nat. Genet. 43, 561–564 (2011).
4. Dang, D. T., Pevsner, J. & Yang, V. W. The biology of the mammalian Kruppel-like family of transcription factors. Int. J. Biochem. Cell Biol. 32, 1103–1121 (2000).
5. Wu, Z. & Wang, S. Role of kruppel-like transcription factors in adipogenesis. Dev. Biol. 373, 235–243 (2013).
6. Parker-Katiares, L. et al. Identification of the imprinted KLF14 transcription factor governing human-specific accelerated evolution. PLoS Genet. 3, e65 (2007).

7. Buil, A. et al. Gene–gene and gene–environment interactions detected by transcriptome sequence analysis in twins. Nat. Genet. 47, 88–91 (2015).

8. Greenawalt, D. M. et al. A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. Genome Res. 21, 1008–1016 (2011).

9. Cvekl, M. et al. Genetic regulation of adipose gene expression and cardio-metabolic traits. Am. J. Hum. Genet. 100, 428–443 (2017).

10. Emilsson, V. et al. Genetics of gene expression and its effect on disease. Nature 452, 423–428 (2008).

11. Keildson, S. et al. Expression of phosphofructokinase in skeletal muscle is influenced by genetic variation and associated with insulin sensitivity. Diabetes 63, 1541–1545 (2014).

12. Innocenti, F. et al. Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. 7, e1002078 (2011).

13. van de Bunt, M. et al. Transcript expression data from human islets links regulatory signals from genome-wide association studies for type 2 diabetes and glycemic traits to their downstream effectors. PLoS Genet. 11, e1005694 (2015).

14. GTEx Consortium. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).

15. DLAG-Genetics Research And Meta-analysis (DIAGRAM) Consortium. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat. Genet. 46, 234–244 (2014).

16. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).

17. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015).

18. McCarthy, D. J. et al. Testing for association using genotypes from a case–control study. Nat. Genet. 43, 1031–1035 (2011).

19. Naranbhai, V. et al. Genomic modulators of gene expression in human islets links regulatory lexicon. Nat. Biotechnol. 33, 555–562 (2015).

20. Rottig, A. et al. Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. PLoS Genet. 7, e1002367 (2011).

21. Fang, I. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol. Cell 49, 359–367 (2013).

22. Rottig, A. et al. Impact of age, BMI and HbA1c levels on the genome-wide genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. 48, 3792–3813 (2015).

23. Guo, Y. et al. Periheltrine activates KLF14 and reduces atherosclerosis by modulating ApoA-1 production. J. Clin. Invest. 125, 3819–3830 (2015).

36. Arner, E. et al. Adipocyte turnover: relevance to human adipose tissue morphology. Diabetes 59, 105–109 (2010).

37. Hammad, N. et al. Adipose tissue dysfunction and reduced insulin sensitivity. Diabetologia 51, 1541–1545 (2014).

38. Ronn, T. et al. Impact of age, BMI and HbA1c levels on the genome-wide genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. 48, 1498–1506 (2000).

39. Almhall, M. et al. Independent and opposite associations of waist and hip circumferences with diabetes, hypertension and dyslipidemia: the AusDiab Study. Int. J. Obs. Relat. Metab. Disord. 28, 402–409 (2004).

40. Yusuf, S. et al. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case–control study. Lancet 366, 1640–1649 (2005).

41. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods 9, 215–216 (2012).

Acknowledgements

This study was supported by MRC grant MR/R010642/1 to K.S.S., R.D.C., F.K. and M.I.M. K.S.S. is supported by an MRC New Investigator Award (MR/L01999X/1). M.I.M. is a Wellcome Senior Investigator and is supported by Wellcome (090532, 106130, 108596, 095101/Z/09/Z), the UK Meta-Obesity Team (MR/L020149/1). F.K. is supported by the British Heart Foundation (RG/17/1/36633). M.C. is supported by NIH Award R00 HL112172. A.G.L. is a Wellcome Senior Fellow in Basic Biomedical Science (095101/10/Z and 200837/Z/16/Z). K.M. is supported by NIH award R01-DK99571, and A.J.L. is supported by NIH award NIH P01HL28481. R.D.C. is supported by MRC MC_UP14661184. J.F. was a Marie Curie Fellow. M.L. is supported by Academy of Finland grants 32799 and 134243, the Finnish Heart Foundation, the Finnish Diabetes Foundation, and Commission of the European Community’s 7th Framework Programme (FP7/2007-2013). The TwinsUK study also receives support from the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust in partnership with King’s College London. We acknowledge excellent technical support for animal husbandry (Mary Lyon Centre), genotyping, histology and pathology. We thank the volunteers from the Oxford Biobank, NIHR Oxford Biomedical Research Centre, for their participation. The Oxford Biobank (http://www.oxfordbiobank.org.uk) is also part of the NIHR National Bioresource, which supported the recalling process of the volunteers. GERDA data came from a grant, the Resource for Genetic Epidemiology Research in Adult Health and Ageing (RC2 AG030667; C. Schafer and N. Risch, principal investigators), awarded to the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics. The RPGEH was supported by grants from the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, and the Ellison Medical Foundation. This research utilized data from the UK Biobank Resource under application 9161.

Author contributions

K.S.S., A.L.G., K.M., A.J.L., R.C., F.K. and M.I.M. conceived of and designed the project; M.T., M.C., J.S.E.-S.M., M.M.S., J.F.-T., C.A.G., L.Q., C.P., P.-C.T., A.N. and G.T. collected and analyzed data; T.M., X.W., A.H., S.S., M.Y., N.C., A.R. and P.A. contributed data; K.S.S., M.T., M.C., J.S.E.-S.M., M.M.S., K.M., A.J.L., R.C., F.K. and M.I.M. wrote the manuscript; and all authors read and approved of the manuscript.

Competing interests

M.I.M. has received consulting and advisory board honoraria from Pfizer, Lilly and Novo Nordisk; and G.T. and U.T. and K.S. are employees of deCODE Genetics/Amgene.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0088-x.

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

Correspondence and requests for materials should be addressed to K.S.S. and M.I.M.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
**Methods**

**Choice of lead SNP.** The T2D association at the KLF14 locus consists of 29 SNPs in near perfect linkage disequilibrium (r^2 > 0.94). The lead SNP in T2D GWAS has been reported as rs4731702 (Kong et al.), rs972283 (Voight et al.) and rs10954284 (Morris et al. and Mahajan et al.), all three of which are in perfect linkage disequilibrium in Europeans (r^2 = 1). The lead SNP for the HDL-C GWAS is rs4731702 (Teslovich et al. and Willer et al.). As rs4731702 is present on all commonly used genotyping arrays, we chose to report all results in this manuscript with respect to rs4731702, where C is the T2D risk allele and T is the non-risk allele.

**TwinUK gene expression.** Biopsies and blood samples from 856 healthy female twins from the TwinsUK cohort were collected within the MuTHER project, and RNA was sequenced as previously described (see Supplementary Note 1 for further details of the TwinsUK data).

**cis- and trans-eQTL analysis.** Cis- and trans-eQTL analysis was conducted in the TwinsUK RNA-seq data as follows. Exon read counts were corrected for technical covariates and family structure using a mixed-effects model, including insert size and mean GC content as fixed effects, and primer index, sample processing date, family and zygosity as random effects. Corrected residuals were used for all eQTL analyses. Cis- and trans-eQTL associations were conducted using the MatrixeQTL package using a standard additive linear model, with BMI, age and age-squared included as covariates. A window size of 1 Mb was used for cis-eQTL analyses. A FDR threshold of 5% was applied to trans-eQTL results, with FDR calculated by quantifying the false discovery rate (FDR).

**Functional fine mapping.** We used the chromatin states predicted by ChromHMM using 127 reference epigenomes from the ENCODE and Roadmap Epigenomics projects. ChromHMM segments the genome into 15 states using 5 chromatin marks (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3) with predictions ranging from active transcription start sites to repressed Polycrom complexes. The ChromHMM predicted enhancer spans a 1.6-kb region, ~4 kb upstream of KLF14 transcription start site (TSS), and harbors five of the T2D GWAS SNPs (rs12154627, rs6973807, rs6974040, rs6974288, rs11762784).

**TwinUK methylation data.** Adipose tissue sample in 603 individuals (taken from the same biopsy as the RNA-seq data) and 309 whole-blood samples from the TwinsUK study were profiled on the Illumina Infinium HumanMethylation450 BeadChip previously. The following covariates were included in analyses: batch (beadchip), BS conversion efficiency (as assessed using the built-in BS conversion efficiency controls) and BS-treated DNA input. The association of methylation was additionally corrected for estimated cell counts. To compare methylation and expression levels, each was first adjusted for covariates, and Pearson’s correlation performed on the residuals.

**Trans-network mediation analysis.** Significant mediation was determined by computing Sobel’s test statistic. To calculate the mediation score, the three following models were implemented:

**Model 1:**

\[ Y = \beta_0 + \beta_1 P + \beta_2 G + e \]

**Model 2:**

\[ Y = \beta_0 + \beta_1 X + \beta_2 A + \beta_3 P + \beta_4 G + e \]

**Model 3:**

\[ \text{Mediation score} = \frac{\beta_1 - \beta_3}{\beta_3} \]

where Y = trans-gene expression, A = age, E = cis-gene expression, P = BMI, and G = cis-genotype.

By conditioning on cis-gene expression (the mediator, E), we can determine if each, individual association detected in trans is regulated in cis or is independent by quantifying \( \Delta H \).

**Enrichment of KLF14 binding in trans genes by Chip-seq analysis.** We used the HEK293 KLF14 Chip-seq data from Najafabadi et al. Genomic coordinates of the peaks of KLF14 binding (along with the summit position for each peak) identified in the Chip-seq experiments were available in the supplementary data of Najafabadi et al. (http://hugheslab.cbr.utoronto.ca/supplementary-data/CH2H_B1H1). For the enrichment analysis for KLF14 binding, we used the summit position of the 18,652 distinct Chip-seq peaks of KLF14 binding (filtered on a minimum Phred-like quality score of call confidence of 50) provided in these data.

An enrichment for KLF14 binding sites for the 385 genes with trans-expression association with the KLF14 locus (FDR < 5%) was evaluated by comparing the proportion of genes for which KLF14 binding sites were identified in the following two groups of genes: first, the 385 genes with trans-expression association with the KLF14 locus; and second, the remaining genes that were tested in the trans-expression analysis (n = 18,678). Enrichment for KLF14 binding sites for the 385 genes with trans-expression association was further evaluated by performing 10,000 permutations of a random set of 385 genes sampled from the entire set of genes that was tested in the trans-expression analysis, and estimating the number of genes with KLF14 binding sites for each permutation.

**iRegulon transcription factor–binding analyses.** In order to find the KLF14 regulon (a transcription factor (TF) and its direct transcriptional targets, which contain common TF binding sites in their cis-regulatory control elements), we used the iRegulon cytoscape plugin. The prediction of regulons consists of four steps. (1) Motif detection: this process relies on an offline scoring step of a sequence search space (10kb or 20kb around the TSS), whereby every gene in the human genome, along with orthologous sequences in ten other vertebrate genomes, is scanned with Cluster-Buster homotopic clusters of motifs using a library of nearly 10,000 position weight matrices (PWMs), resulting in a gene ranking list for each PWM. Enriched motifs are those motifs for which the input genes are enriched at the top of the ranking, using the area under the curve (AUC) of the cumulative curve. (2) Track discovery: this step also relies on an offline scoring step whereby every gene in the human genome is scored with around 1120 Chip-seq tracks, generating a list of TFs according to the highest ChIP peak within the regulatory space. (3) Motif2TF mapping: the candidate TFs are prioritized by finding the optimal path from a motif to a TF, in a motif-TF network. (4) Target detection: for each enriched motif, the candidate targets are selected as the significantly highly ranked genes compared to the genomic background and to the entire motif collection as background.

**Functional enrichment analysis.** KLF14 trans genes were subdivided into three subsets on the basis of the presence of bindings motifs for KLF14, SREBF1 or neither. The three subsets were assessed for functional enrichment using TopFun, part of the ToppGene package. Results were corrected for multiple testing using the Bonferroni method, with enrichments considered significant at a Bonferroni-corrected P value of <0.05.

**Overlap between KLF14 trans-regulated genes and published GWAS results.** The NHGRI-EBI GWAS catalog (available at http://www.ebi.ac.uk/gwas, accessed November 2015) was filtered so as to retain signals meeting genome-wide significance (\( P < 5.0 \times 10^{-8} \)). Corresponding cis-eQTLs for each of the 385 KLF14 trans genes, including secondary, tertiary and quaternary cis-eQTLs sequentially conditioned on the lead cis-eQTLs at each locus (a total of 465 cis-eQTLs), were then cross-referenced against the NHGRI GWAS results, retaining at each KLF14 trans-gene locus those reported NHGRI associations within ±250kb of the relevant cis-eQTLs that showed moderate to high LD (\( r^2 \geq 0.8 \)) between the reported GWAS SNP and the respective cis-eQTL.

**Association between KLF14 expression and TwinsUK metabolic traits.** To determine if gene expression was associated to a concurrently measured phenotype of interest, expression of each exon was treated as a quantitative trait in linear mixed effects model implemented using the lme4 package. The linear mixed-effects model was adjusted for age, BMI and technical covariates (mean GC content and insert size mode) (fixed effects), family relationship (twin pairing), primer index and zygosity (random effects). Phenotypic data were treated as continuous independent variables and were inverse-normal-transformed. A full model with the phenotype fitted was compared to a null model (no phenotype) using a 1-df ANOVA. An FDR was estimated using the package QVALUE to obtain \( q \) values that correspond to a controlled FDR 5%. TwinsUK metabolic phenotypes were measured at the same time-point as the biopsy and were collected as previously described. Body-fat distribution traits were measured using dual-energy X-ray absorptiometry (DXA; Hologic QDR 4500 plus) with the standard protocol.

**Trait-expression association enrichment analyses.** The most significantly associated exon per gene was retained for all genes. Enrichment was assessed by comparing the number of genes significantly associated with each phenotype (at \( q < 0.05 \)) within the set of KLF14 trans genes \((n_{trans-gene} = 385)\) to the remaining genes not within the KLF14 trans-network \((n_{trans} = 18,716)\), using a Fisher’s exact test.

**Sex x SNP interaction analysis in large cohorts.** Details of the data preparation, quality control and cohort specific analysis covariates for the UK Biobank, GERA and WTCCC/T2D data are provided in Supplementary Note 1. To evaluate the SNP x sex interaction effect of rs4731702 on anthropometric traits in the UK Biobank, we tested for association of the genotype with each inverse-normalized variable using linear regression with a main effect for SNP and SNP-by-sex interaction terms using an additive model. We adjusted for covariates: age, sex, six (within UK) ancestry principal components, and array version used to
Articles Nature Genetics

Klf14

test strips (OneTouch). For histological examination, fat tissue was fixed in 4%

weight. Whole-blood glucose levels were measured using a glucose meter and

60 and 120 min after intraperitoneal injection of glucose (1 g/kg body weight).

Diagnostics) independently according to the manufacturers' instructions. For the

following an overnight fast. Samples were kept on ice then centrifuged for 10 min

permitted by the restrictions set forth by institutional policy (typically no more

Animal Care and Use Committee and were consistent with local, state, and federal

models.

Experimental design of in vivo mouse experiments in CRISPR-Cas9 mouse models.

The CRISPR-Cas9 knockout mouse models were generated as described in Supplementary Note 1. All procedures used for CRISPR-Cas9 animal studies were approved by Harvard University's Faculty of Arts and Sciences Institutional Animal Care and Use Committee (IACUC) protocol 14-05-202. Klf14tm1(KOMP)Vlcg mice were kept and studied in accordance with UK Home Office legislation and local ethical guidelines issued by the Medical Research Council (Responsibility in the Use of Animals for Medical Research, July 1993; home office license 30/3146).

Experimental design of in vivo mouse experiments in CRISPR-Cas9 mouse models. Klf14tm1(KOMP)Vlcg deletion mice were generated as described in Supplementary Note 1. Deletion mice were generated under controlled conditions of light (7am–7 pm, dark 7pm–7am), temperature (21°C ± 1°C), and humidity (50% ± 10%). They had free access to water (9–13 ppm chlorine) and were fed a commercial diet (Purina Lab 

Note 1. Deletion mice were kept under controlled conditions of light (light

7am–7 pm, dark 7pm–7am), temperature (21°C ± 1°C), and humidity (50% ± 10%). They had free access to water (9–13 ppm chlorine) and were fed a commercial diet (Purina Lab

Experimental design of in vivo mouse experiments in deletion mouse models. Klf14tm1(KOMP)Vlcg deletion mice were generated as described in Supplementary Note 1. Deletion mice were generated under controlled conditions of light (light

7am–7 pm, dark 7pm–7am), temperature (21°C ± 1°C), and humidity (50% ± 10%). They had free access to water (9–13 ppm chlorine) and were fed a commercial diet (Purina Lab

Cohorts of male and female mice were bred for longitudinal blood- and body-composition-based phenotyping tests. These included four groups: (1) MAT (heterozygotes (hets) inheriting the allele maternally), (2) PAT (hets inheriting the allele paternally), (3) WT MAT controls, and (4) WT PAT controls. Sample-size estimates were based on previous experience and data from other mouse models in which the relevant traits were measured. These data were used in power calculations to help determine cohort sizes and variances. The genotypes of the following mixed genotype across multiple litters and were not randomized into groups. Mouse IDs and genotypes were recorded on the cages, and the operator carrying out the animal procedure were not blinded to these, although subsequent tests only include animal ID information and clinical chemistry was carried out by a core service without knowledge of the genotypes.

Echo-MRI analysis and fasted blood sampling from the lateral tail vein (local anesthetic) were carried out at 8, 12, 22 and 27 weeks of age. Whole-blood glucose was measured using an AlphaTRAK meter and test strips (Abbott). Plasma insulin was assayed using a Mouse insulin ELISA kit (Mercodia). Terminal blood samples were collected from 33-week-old mice by intraperitoneal anesthesia and cardiac puncture following an overnight fast. Samples were kept on ice then centrifuged for 10 min at 8,000 g in a centrifuge set to room temperature. The resulting plasma was analyzed on a Beckman Coulter AU680 clinical chemistry analyzer using reagents and settings recommended by the manufacturer. Clinical biochemistry for cholesterol analysis was conducted on 8-, 22-, 27- and 33-week blood samples.

Statistical analysis of mouse data. Data collection, summary calculations and descriptive statistics were carried out using Microsoft Excel 2010. Statistical analysis was carried out using software Graphpad Prism v6. Effects of genotype at different time-points on clinical chemistry parameters were determined by two-way ANOVA with repeated measures and Bonferroni's correction for multiple comparisons. Unpaired two-tailed t-tests were used for pairwise comparisons as appropriate and variances were equal, except for Fig. 4a male data, in which a Welch's correction was applied for unequal variance. Datasets were tested for normal distribution and equal variance according to an D'Agostino-Pearson omnibus K2 normality test and a Brown-Forsythe test, respectively, in Graphpad Prism. Consequently, non-parametric tests were applied as necessary and are indicated in the figures. In these cases under the curve was calculated (baselined to t = 0 values) and tested with either a one-way ANOVA Kruskal-Wallis test and Dunns multiple comparison test where there were multiple groups (Klf14tm1(KOMP)Vlcg knockout) or Mann-Whitney two-tailed t-test when there were only two groups (CRISPR-Cas9 knockout). Furthermore, as indicated, individual pairwise comparisons were made using Mann-Whitney two-tailed t-tests and are shown in the figures.

Gene expression in mouse models. Mouse RNA-seq data were collected as described in Supplementary Note 1 and were analyzed by the in-house bioinformatics team at MRC Harwell using a previously described pipeline24. Functional enrichment for mouse differentially expressed genes was assessed using ToppFun24, with enrichment considered significant after FDR multiple-testing correction (q < 0.05). MEME-chip was used to identify motifs that were similarly over-represented in DNase I-hypersensitivity hotspots upstream of mouse differentiated genes.

Isolation, culture and differentiation of pre-adipocytes. Abdominal subcutaneous adipose tissue (ASAT) biopsies were obtained from eight healthy, surgically-examined adult females and one male (12 per case).

For lipid measurements, blood samples were collected from the lateral tail vein following an overnight fast. Samples were kept on ice then centrifuged for 10 min at 2,000 g in a centrifuge at 4°C. The plasma levels of triglyceride, total cholesterol and HDL-C were measured using Infinity Triglycerides Reagent (Thermo Fisher), Infinity Cholesterol Reagent (Thermo Fisher) and HDL-Cholesterol E (Wako Diagnostics) independently according to the manufacturers’ instructions. For the intraperitoneal glucose tolerance test (IPGTT), the mice were fasted overnight, and glucose concentrations were measured immediately before and at 20, 40, 60 and 120 min after intraperitoneal injection of glucose (1 g/kg body weight). For the insulin tolerance test (ITT), the mice were fasted for 4 h, and glucose concentrations were measured immediately before and at 20, 40, 60 and 120 min after intraperitoneal injection with human insulin (Sigma) at 1 U/kg body weight. Whole-blood glucose levels were measured using a glucose meter and test strips (OneTouch). For histological examination, fat tissue was fixed in 4% paraformaldehyde and embedded in paraffin. Consecutive sections were cut and stained by hematoxylin and eosin for light microscopy examination and evaluation.

Isolation, culture and differentiation of pre-adipocytes. Total RNA was extracted from pre-adipocytes as described25. Real-time PCR was performed on an Applied Biosystems 7900HT, using TaqMan Assays-on-Demand (Applied Biosystems) and Klear Master Mix (KBiosciences). mRNA expression values for target genes were calculated using the ΔCt transformation method26. The ΔCt was calculated as follows: ΔCt = efficiency (minimum of sample Ct). Values were normalized to energy control genes (PPIA and UBC)

Explant lipogenesis experiments. Adipocytes isolated from abdominal subcutaneous adipose tissue were used to measure activation of insulin-stimulated lipogenesis, as described in detail elsewhere27. In brief, isolated human adipocytes were incubated at a concentration of 2% (w/v) in Krebs–Ringer phosphate buffer (pH 7.4) containing albumin (40 mg/ml), [3–3H] glucose (5 x 10^-4 dpm/ml), unlabelled glucose (1 μmol/l) and varying concentrations of human insulin (0–70 nmol/l). The incubations were conducted for 2h at 37°C with air as the gas phase. Incubations were stopped by rapidly chilling the incubation vials to 4°C, and the amount of glucose incorporated into adipose tissue (insulin-stimulated lipogenesis) was determined. Lipogenesis was expressed as the amount of glucose incorporated either per lipid weight of fat cells or per fat cell number, as described previously27.

Short hairpin RNA-mediated silencing of KLF14. KLF14 was silenced in primary pre-adipocytes derived from female ASAT. Lentiviral particles were produced by transient co-transfection of HEK293 cells, using a KLF14-specific short hairpin RNA (shRNA) lentiviral transduction construct (SHCLNV-NM_138693.1-255151c1; Sigma-Aldrich) and non-targeting shRNA lentiviral constructs (SHC002V; Sigma-Aldrich) with packaging vectors (MISSION; Sigma-Aldrich). Cells were stably transfected by transduction of pre-adipocytes with lentiviral particles followed by selection in growth media containing 2 mg/ml puromycin.

Quantification of intracellular lipid content. Control and KLF14-specific shRNA cell lysates were prepared in lysis buffer containing 1% Igepal-630, 150 mM NaCl and 50 mM Tris HCl. Lysates were sonicated and an aliquot was collected for protein quantification using the BioRad DC Protein Assay kit. The remaining
lysat was heated at 95 °C for 30 min, allowed to cool to room temperature, and then centrifuged at 12,000 g for 10 min. The triglyceride concentration in the supernatant was determined using a commercially available enzymatic kit (Instrumentation Laboratory UK) on an iLAB 650 Chemistry Analyser. Total triglycerides were normalized to protein concentration.

**Estimation of cell doubling time.** Equal numbers of control and KLF14-specific shRNA–specific adipocytes were seeded in T75 flasks. 1.5 × 10^6 cells. Cells were trypsinized and double counted every 5 days. Doubling time was calculated using the formula

\[ T_d = (\ln (2)/\ln(Q_2/Q_1)) \times \frac{Q_2 - Q_1}{Q_2} \]

where \( T_d \) = time (days) and \( q = \) cell number.

**Insulin-stimulated glucose uptake.** Glucose uptake was assayed according to the established protocol from a commercial glucose uptake kit (1342; Promega). The luminescent glucose uptake assay was applied to control and KLF14-specific shRNA–differentiated adipocytes in 96-well plates. In brief, differentiated cells were starved in hormone-free DMEM/F12 medium overnight. The medium was removed and the cells were washed with 100 µl of phosphate-buffered saline (PBS), followed by incubation with 10 mM insulin for 1 h at 37 °C in 5% CO2. To initiate glucose uptake, 50 µl of 2DG (1 mM) in PBS was added to the cells. The uptake reaction was stopped and the samples were processed as described in the manufacturer’s protocol (1342, Promega). All assay steps were performed at room temperature. All data were acquired on a PerkinElmer EnSpire 2300 multimode plate reader instrument, with an integration time of 0.5 s.

**Adipocyte cell size and number study population and sample collection.** Nine females and nine males carrying KLF14 risk allele and paired control subjects matched for age (30–50 years) and BMI (22–27 kg/m2) were recruited from the OBB. Paired abdominal and gluteal subcutaneous adipose specimens were obtained by gun and needle biopsies. The gun biopsies were fixed in 10% paraformaldehyde, embedded in paraffin wax, cut into 5-µm sections and stained with hematoxylin and eosin. Sections were viewed at 20× magnification, and adipocyte cross-sectional area was calculated using Adobe Photoshop 5.0.1 (Adobe Systems) and Image Processing Tool Kit (Reindeer Games). As previously described, in order to accurately determine the minimum number of cells required for measurement of cell size distribution in a sample, we took four samples and counted 1,000 cells in each and observed that the coefficient of variation started to increase when fewer than 100 cells were included in each biopsy. Therefore, we included only biopsies with more than 100 cells available for quantification (n = 9 pairs from each genotype). Statistical significance was assessed using a Wilcoxon signed-rank test. Replication of the cell size was undertaken in a Swedish cohort. Abdominal subcutaneous adipose specimens were obtained by needle biopsy. Adipocytes were separated from stroma cells by treatment in a shaking bath at 37 °C for 60 min with collagenase (0.5 mg/l) in 5 ml Krebs Ringer phosphate buffer (pH 7.4) with purified BSA (40 g/l) as previously described. Adipocyte suspensions were then rinsed three times in collagenase-free buffer using nylon filters, and the cell sizes were measured by direct microscopy. The mean adipocyte diameter was calculated from measurements of 100 cells, and statistical significance assessed using a Wilcoxon signed-rank test. Cell numbers were calculated using the formula described, where DXA-measured adipose depot size was divided by mean cell volume.

**Life Sciences Reporting Summary.** Further information on experimental design is available in the Life Sciences Reporting Summary.

**Data availability statement.** TwinsUK RNA-seq data are available from EGA (EGAS00001000805). TwinsUK adipose methylation data are available from ArrayExpress (E-MTAB-1866), and blood methylation data are available from GEO (GSE50660). TwinsUK genotypes are available upon application to the TwinsUK cohort. METSIM adipose array data are available from GEO (GSE70353).

### References

44. Grundberg, E. et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. 44, 1084–1089 (2012).
45. Shabalina, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 28, 1353–1358 (2012).
46. Stroyer, J. D. & Tibshirani, R. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA 100, 9440–9445 (2003).
47. Stroyer, J. D., Bass, A. J., Dabney, A. & Robinson, D. qvalue: Q-value estimation for false discovery rate control. R package version 2.6.0 http://github.com/jdstorey/qvalue (2015).
48. Grundberg, E. et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am. J. Hum. Genet. 93, 876–890 (2013).
49. Tsaprouni, L. G. et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Epigenetics 9, 1382–1396 (2014).
50. Pierce, B. L. et al. Mediation analysis demonstrates that trans-eQTLs are often explained by cis-mediation: a genome-wide analysis among 1,800 South Asians. PLoS Genet. 10, e1004818 (2014).
51. Najafabadi, H. S., Abzu, M. & Hughes, T. R. Identification of C2H2-ZF binding preferences from ChIP-seq data using RCAde. Bioinformatics 31, 2879–2881 (2015).
52. Janky, R. et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. PLoS Comput. Biol. 10, e1003731 (2014).
53. Frith, M. C., Li, M. C. & Weng, Z. Cluster-Buster: finding dense clusters of motifs in DNA sequences. Nucleic Acids Res. 31, 3666–3668 (2003).
54. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP–trait associations. Nucleic Acids Res. 42, D1001–D1006 (2014).
55. Parsons, M. J. et al. The regulatory factor ZFHX3 modifies circadian function in SCN via an AT motif-driven axis. Cell 162, 607–621 (2015).
56. Collins, J. M., Neville, M. J., Hoppa, M. B. & Frayn, K. N. De novo lipogenesis and stearoyl-CoA desaturase are coordinately regulated in the muscle via an AT motif-driven axis. Nucleic Acids Res. 31, e45 (2003).
57. Frith, M. C., Li, M. C. & Weng, Z. Cluster-Buster: finding dense clusters of motifs in DNA sequences. Nucleic Acids Res. 31, 3666–3668 (2003).
58. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP–trait associations. Nucleic Acids Res. 42, D1001–D1006 (2014).
59. Parsons, M. J. et al. The regulatory factor ZFHX3 modifies circadian function in SCN via an AT motif-driven axis. Cell 162, 607–621 (2015).
60. Collins, J. M., Neville, M. J., Hoppa, M. B. & Frayn, K. N. De novo lipogenesis and stearoyl-CoA desaturase are coordinately regulated in the muscle via an AT motif-driven axis. Nucleic Acids Res. 31, e45 (2003).
61. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res. 29, e45 (2001).
62. Neville, M. J., Collins, J. M., Glyn, A. L., McCarthy, M. I. & Karpe, F. Comprehensive human adipose tissue mRNA and microRNA endogenous control selection for quantitative real-time-PCR normalization. Obesity 19, 888–892 (2011).
63. Hofstredt, J., Naslund, E., Wren, M. & Arner, P. Prospective and controlled studies of the actions of insulin and catecholamine in fat cells of obese women following weight reduction. Diabetologia 48, 2334–2342 (2005).
64. Marinou, K. et al. Structural and functional properties of deep abdominal subcutaneous adipose tissue explain its association with insulin resistance and cardiovascular risk in men. Diabetes Care 37, 821–829 (2014).
65. Dahlman, I. et al. Numerous genes in loci associated with body fat distribution are linked to adipose function. Diabetes 65, 433–437 (2016).
66. Hirsch, J. & Gallian, E. Methods for the determination of adipose cell size in man and animals. J. Lipid Res. 9, 110–119 (1968).
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study. For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

1. Sample size
   - Describe how sample size was determined.

   Human transcriptomic studies: Transcriptomic data was from previously published studies, the sample size was determined by the availability of the published data.

   Human Cellular studies: We assumed group mean (matched carrier (genotype A) vs non-carriers (genotype B)) with an estimated phenotypic difference of 30% (μ_A-μ_B) where the SD is 0.2 of μ. This scenario would require 7 people in each group to secure 80% power to show a statistically significant difference (p<0.05). The observed difference for adipose tissue cell size was almost 50% and highly statistically significant.

   Mouse Models: Sample size estimates were based on previous experience and data from other mouse models in which the relevant traits were measured. These data were used in power calculations to help in deciding cohort sizes.

2. Data exclusions
   - Describe any data exclusions.

   Human transcriptomic studies: Post-menopausal women taking hormone replacement therapy were excluded from the expression vs menopausal status analysis as the effect of hormone replacement therapy can mimic pre-menopausal phenotypes. Individuals who did not fast before data collection were excluded from expression vs metabolic trait analysis.

   Human Cellular studies: Adipose biopsies with less than 100 cells were not included in the analysis of cell size. This threshold was determined by taking 4 samples and counting 1,000 cells in each. The data were then removed by 100 at a time and it was observed that the coefficient of variation started to increase when less than 100 cells were included in each biopsy.

3. Replication
   - Describe the measures taken to verify the reproducibility of the experimental findings.

   All experimental observations/findings were successfully replicated and results reliably reproduced.

4. Randomization
   - Describe how samples/organisms/participants were allocated into experimental groups.

   Human cellular studies: Human cellular studies: for human studies subjects were selected by genotype from the Oxford Biobank and then recruited for tissue biopsies. Although the recruitment into the genotypic groups involved matching for age and BMI, final statistical analysis comparing the genotypic groups was also matched for age and BMI.

   Mouse Models: Deletion mice were were housed in single sex groups of mixed genotype across multiple litters and were not randomised into groups. Crispr-CAS-9 control mice and KO mice were littermates from several litters that were timed to be the same age; the mice were kept in the same cages, in as few cages as permitted by the restrictions set forth by institutional policy (typically no more than three adult mice per cage).

5. Blinding
   - Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Human cellular studies: investigators were blinded to group allocation during data collection and analysis.

   Mouse models: Mouse IDs and genotypes were recorded on the cages and were not blinded to the operator carrying out the animal procedure although subsequent tests only include
animal ID information and clinical chemistry was carried out by a core service without knowledge of the genotypes.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

- The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- Test values indicating whether an effect is present
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

7. Software

Describe the software used to analyze the data in this study.

Matrix-eQTL; iREgulon, ToppGENE; Qvalue v 1.4 Graphpad Prism v6; MEME; SHAPEIT v2.5;
PLINK; IMPUTE2; Cluster-Buster; R v3.1.1; BM SPSS Statistics V22.0

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

- Not applicable

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

- Not applicable

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

- Not applicable

b. Describe the method of cell line authentication used.

- Not applicable

c. Report whether the cell lines were tested for mycoplasma contamination.

- All primary cells were tested and were negative for mycoplasma

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

- Not applicable
Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines.

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

Deletion mouse model: The mouse strain (Klf14tm1(KOMP)Vlcg) used in this project was generated from targeted embryonic stem (ES) cells for Klf14 obtained from the KOMP repository www.komp.org, a NCRR-NIH supported mouse repository. The U42-RR024244 ES cells from which this mouse was generated were created by Velocigene from funds provided by the trans-HIH Knock-out Mouse Project (KOMP) (Grant number SU0101HG004085). Live mice were imported on a C57BL/6NTac USA background and rederived into the MRC Harwell Mary Lyon Centre specified pathogen free (SPF) facility and maintained on C57BL/6NTac. Mice were kept and studied in accordance with UK Home Office legislation and local ethical guidelines issued by the Medical Research Council (Responsibility in the Use of Animals for Medical Research, July 1993; home office license 30/3146). Mice were kept under controlled light (light 7am–7pm, dark 7pm–7am), temperature (21±2°C) and humidity (55±10%) conditions. They had free access to water (9–13 ppm chlorine) and were fed ad libitum on a commercial diet (SDS Rat and Mouse No. 3 Breeding diet, RM3) until 18 weeks of age when they were then transferred to a high fat diet (45% kcal from fat; D12451; Research Diets). Equal numbers of male and female mice were included.

Crispr-Cas9 mouse models: All procedures used for CRISPR-Cas9 animal studies were approved by Harvard University’s Faculty of Arts and Sciences Institutional Animal Care and Use Committee and were consistent with local, state, and federal regulations as applicable. All animals were housed in a 12-hour dark-light cycle and fed standard chow ad libitum in conditions similar to those described above. Mice were derived on a C57BL/6J background. Equal numbers of male and female mice were included.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

This study recruited 36 individuals from the Oxford BioBank (OBB) in Oxford, United Kingdom. Individuals were within the age range 30-50 and BMI range 22-27. Study participants were half female and half male, and within each sex an equal number of KLF14 risk allele homozygotes and age and BMI matched non-risk allele homozygotes were recruited. The study was approved by Oxfordshire Clinical Research Ethics Committee (08/H0606/107+S) and all subjects gave written informed consent.

Human transcriptomic data was taken from published studies. Relevant population characteristics are included in Supplementary Table 2.