Genome-wide association studies in obesity have identified a large number of non-coding loci located near genes expressed in the central nervous system. However, due to the difficulties in isolating and characterizing specific neuronal subpopulations, few obesity-associated single-nucleotide polymorphisms have been functionally characterized. Leptin-responsive neurons in the hypothalamus are essential in controlling energy homeostasis and body weight. Here, we combine fluorescence-activated cell sorting of leptin-responsive hypothalamic neuron nuclei with genomic and epigenomic approaches (RNA sequencing, chromatin immunoprecipitation sequencing, assay for transposase-accessible chromatin sequencing) to generate a comprehensive map of leptin response-specific regulatory elements, several of which overlap obesity-associated genome-wide association study variants. We demonstrate the usefulness of our leptin response neuron regulome, by functionally characterizing an enhancer near Socs3, a leptin response-associated transcription factor. We envision our data to serve as a useful resource and a blueprint for functionally characterizing obesity-associated single-nucleotide polymorphisms in the hypothalamus.

Results

Nuclei isolation from leptin-responsive neurons. The hypothalamus is a complex brain structure composed of neuronal and non-neuronal cells such as astrocytes, microglia, oligodendrocytes and endothelial cells. This complexity also makes separating these cells from one another for subsequent genome analyses extremely complex. To map the genes and regulatory elements that are active in leptin-responsive neurons (a subset of hypothalamic neurons that includes pro-opiomelanocortin (Pomc), agouti-related protein (Agrp) and neuropeptide Y (Npy) neurons), we generated mice in which the nuclei of these cells were fluorescently labelled and used these nuclei for subsequent cell isolation. Specifically, we crossed Leprb<sup>cre</sup> mice, which carry an internal ribosome entry site followed by Cre recombinase downstream of exon 18b of the leptin receptor (Lepr) gene, with mice that express, in a Cre-dependent manner, the SUN-domain-containing protein 1 (SUN1) nuclear membrane protein fused to green fluorescent protein (GFP) (Fig. 1a). We confirmed that GFP labelling was restricted to the nuclei of these cells by immunofluorescent staining (Supplementary Fig. 1). We then established a protocol to isolate the cell nuclei from the dissected hypothalamus (see Methods). This was followed by fluorescence-activated cell sorting (FACS) to separate the GFP<sup>+</sup> from the GFP<sup>−</sup> nuclei (Fig. 1c and Supplementary Fig. 1).
As the nutritional and/or leptin-activation status may affect the transcriptome or regulome of these neurons, we performed the nuclear isolation under different conditions. Groups of mice were fed, fasted or (to isolate the specific effect of leptin) given leptin injections chronically or acutely during fasting; experiments were carried out at the same time points during the day (Fig. 1d). For each condition, the isolation of nuclei was done from the hypothalamus and pooled from both male and female (to account for sex-specific differences) adult mice (2–4 months old) (Supplementary Table 1). For each assay, at least 5 mice per RNA sequencing (RNA-seq) or assay for transposase-accessible chromatin sequencing (ATAC-seq) and at least 20 mice for chromatin immunoprecipitation sequencing (ChIP-seq) with three biological replicates for RNA-seq (total of 60 mice) and two biological replicates for ATAC-seq (total of 40 mice) and ChIP-seq (total of 160 mice) were used.

**Nuclear transcriptome of leptin-responsive neurons.** To validate enrichment for the nuclei of LepRb-expressing neurons, before mapping the regulatory elements using H3K27ac ChIP-seq and ATAC-seq, and to provide a solid basis for matching transcriptional status with gene regulatory status in the nuclei from leptin-responsive neurons, we analysed our RNA-seq data set for long non-coding RNAs (lncRNAs) that are differentially expressed, as compared with genes. We detected significantly higher levels of expression in the ME3 and a variety of terms for ME6 (Supplementary Fig. 3b). Two of the most significantly correlated modules, ME3 and ME6, showed an enrichment for genes involved in neuronal activity for two of the most significantly correlated modules, ME3 and ME6, and as such eating behaviour and regulation of appetite (Supplementary Fig. 3b).

**Fig. 1 | Isolation of LepRb-GFP+ nuclei. a.** R26-CAG-LSL-Sun1-sfGFP-myc homozygous mice were crossed with Leprb<sup>tm6</sup> mice to obtain mice that express sfGFP in LepRb-expressing cell nuclei. IRES, internal ribosome entry site. b. Immunostaining showing sfGFP expression in the DMH, VMH and ARC nucleus. Right panel: zoom in on the ARC along with DAPI staining (blue). c. FACS histogram showing clear separation of GFP<sup>+</sup> from GFP<sup>−</sup> nuclei obtained from Leprb<sup>tm6</sup> x R26-CAG-LSL-Sun1-sfGFP-myc heterozygous mice. Similar results were obtained in all the experimental groups (Supplementary Fig. 1). d. Leprb<sup>tm6</sup> x R26-CAG-LSL-Sun1-sfGFP-myc heterozygous mice were fed or fasted for 36 h, during which saline (blue) or leptin (orange) was injected at various time points. Cell nuclei were isolated and sorted via FACS, followed by RNA-seq, ChIP-seq and ATAC-seq.

Importantly, unlike translatome analysis, nuclear transcriptome analysis provides information on non-translated RNA species. Therefore, we analysed our RNA-seq data set for long non-coding RNAs (lncRNAs) that are differentially expressed, as compared with their expression in all other cell types (GFP−). We found numerous unique lncRNAs and 79 differentially expressed lncRNAs (33 increased and 46 decreased) (Fig. 2c and Supplementary Data 1).
To identify genes that specifically respond to physiological changes in leptin-responsive neurons, we compared the nuclear transcriptomes under different physiological conditions. Analyses of all conditions for our GFP+ nuclei revealed 1,257 genes to be upregulated by leptin and 853 to be downregulated when fasted mice were compared with mice injected every 12 h (Fig. 3a). Gene ontology analyses indicated enrichment for genes associated with neuronal function (Supplementary Fig. 2b). As expected, we found that the expression of suppressor of cytokine signalling 3 (Socs3), a known leptin-responsive transcriptional target, was decreased that the expression of suppressor of cytokine signalling 3 (Socs3), another gene with known function in tumour cells, but has yet to be associated with leptin response. Another leptin response gene was guanylate cyclase 2C (Gucy2c) (Fig. 3b). Similarly to Socs3, Gucy2c is also a known transcriptional target of Stat3, and is known to inhibit its function in tumour cells, but has yet to be associated with leptin response. Another leptin response gene was guanylate cyclase 2C (Gucy2c) (Fig. 3b). Gucy2c encodes a transmembrane receptor that is expressed in the hypothalamus and causes obesity with an increase in leptin serum levels when it is deleted in mice. Combined, our results show that transcriptional analysis of FACS-sorted nuclei can be used to detect transcriptional changes in small neuronal subpopulations under different physiological conditions.

Mapping gene regulatory regions in neuronal subpopulations. To identify the gene regulatory elements that are active in Lepr-expressing neurons, we performed ATAC-seq and ChIP-seq for H3K27ac in sorted nuclei from these neurons. Although the starting material for ATAC-seq is nuclei, thus making the procedure straightforward, performing ChIP-seq required the development of a new protocol after nuclei sorting. Briefly, nuclei were first sorted and then isolated via centrifugation in 1 M CaCl2 and 1 M MgAc2. Gentle fixation using 1% formaldehyde followed by quenching with 125 mM glycine was carried out on the pellet, which was subsequently lysed and subjected to ChIP-seq (see Methods for more details).

We identified a total of 32,377 regions that are active in leptin-responsive neurons, by using ATAC-seq, and 50,218 regions that are active, by using H3K27ac ChIP-seq, respectively. To confirm the relevance and specificity of the identified elements, we compared our RNA-seq results to our ChIP-seq and ATAC-seq results and found that active regulatory regions correlated to levels of gene transcription. Genes near an ATAC-seq or ChIP-seq LepRb-enriched peak in leptin-responsive neurons were significantly more likely to show the same direction of gene expression changes when compared to peaks with non-significant differences between GFP+ and GFP− neurons (P < 9.3 × 10−7 or 2.2 × 10−14 for ATAC-seq and ChIP-seq, respectively; Wilcoxon rank-sum test) (Supplementary Fig. 5). We further examined known leptin-pathway-associated genes and compared the active regions for these genes in leptin-responsive neurons (from GFP+ nuclei after sorting) to those from all the other
hypothalamic cell types (GFP$^+$ nuclei after sorting). As specific examples, we observed a strong enrichment of H3K27ac ChIP-seq and ATAC-seq signals specifically in GFP$^+$ nuclei near the Lep$^{pr}$ gene promoter, the top positively differentially expressed gene in our RNA-seq, as compared with all other cell types (Fig. 2b), and a reduction near the Sim1 promoter (Fig. 4a,b), one of the top negatively differentially expressed genes (Fig. 2b).

We then compared peak enrichment between LepRb-GFP$^+$ nuclei and all other hypothalamic cell types (GFP$^-$). We found 17,629 H3K27ac ChIP-seq peaks to be differentially enriched (Supplementary Data 2), with 9,903 H3K27ac peaks upregulated in GFP$^+$ nuclei and 7,726 downregulated (Fig. 4c,d and Supplementary Data 2). To validate that we were obtaining ChIP-seq peaks that were unique to LepRb-GFP$^+$ nuclei, we next used the Genomic Regions Enrichment ofAnnotations Tool (GREAT27). This tool analyses biologically independent replicates.

To further validate whether our LepRb neuron nuclei regulome contained relevant regulatory regions, we set out to test candidate enhancer sequences for their function. We selected three candidate enhancers located downstream of Socs3 (Socs3-1, Socs3-2 and Socs3-3). These regions were chosen because of their increased enrichment of H3K27ac marks and chromatin accessibility in nuclei of LepRb-expressing neurons (Fig. 5a). We tested the ability of these enhancers to increase the activity of a minimal promoter driving the expression of luciferase in mHypoA-Pomc cells, an established cell line from mouse hypothalamus Pomc-expressing neurons. From the three sequences, one sequence, Socs3-3, showed significant enhancer activity in this cell type (Fig. 5b).

Functional characterization of a Socs3 enhancer. To further validate whether our LepRb neuron nuclei regulome contained relevant regulatory regions, we set out to test candidate enhancer sequences for their function. We selected three candidate enhancers located downstream of Socs3 (Socs3-1, Socs3-2 and Socs3-3). These regions were chosen because of their increased enrichment of H3K27ac marks and chromatin accessibility in nuclei of LepRb-expressing neurons (Fig. 5a). We tested the ability of these enhancers to increase the activity of a minimal promoter driving the expression of luciferase in mHypoA-Pomc cells, an established cell line from mouse hypothalamus Pomc-expressing neurons. From the three sequences, one sequence, Socs3-3, showed significant enhancer activity in this cell type (Fig. 5b).
sgRNA (Fig. 5c). A similar analysis of the neighbouring genes around Socs3, phosphatidylglycerophosphate synthase 1 (Pgs1) and threonine aldolase 1 (Tha1), did not show any significant change in their expression (Fig. 5d,e). Combined, our results confirm that our approach can be used to identify functionally relevant regulatory regions in subpopulations of neurons.

LepRb neuron regulatory elements overlap obesity-associated SNPs. We next set out to test whether obesity-associated GWAS SNPs could be physically linked to the regulatory elements identified in LepRb neurons. We obtained a list of GWAS-reported, obesity-associated SNPs from Ghosh and Bouchard12 and used PLINK31 to obtain a list of all SNPs that are in linkage disequilibrium with...
Fig. 5 | Functional characterization of a Socs3 enhancer. a, Genomic snapshot of the Socs3 locus showing three candidate enhancer sequences, Socs3-1 (1), Socs3-2 (2) and Socs3-3 (3), which are enriched for H3K27ac ChIP-seq and ATAC-seq signals in LepRb-GFP+ nuclei (green) compared to GFP− nuclei (grey). Each of the two replicates for ChIP-seq and ATAC-seq presented comparable signals, as shown. b, Luciferase assay for all three candidate enhancer sequences (Socs3-1, Socs3-2 and Socs3-3). Fold activity is compared to that of the empty vector (pGL4.23) and data are presented as the mean ± the lower and upper quartile; the lines represent the minimum and maximum from three independent biological experiments. c–e, CRISPRi targeting of Soc3-2. Two sgRNAs (Socs3_sgRNA_1 and Socs3_sgRNA_2) that target the Soc3-2 enhancer along with dCas9-KRAB show significantly reduced expression of Soc3 via RT-qPCR when compared to a negative control sgRNA (NC-sgRNA) or a no-infection control (no-sgRNA). The neighbouring genes Pgs1 (d) and Tha1 (e) do not show any significant changes in gene expression. The boxes represent the twenty-fifth to seventy-fifth percentiles; the midline indicates the median. n = 3 biologically independent replicates. A two-sided Student’s t-test was used to determine statistical significance (**P < 0.01; *P < 0.05).

Discussion

Numerous GWAS have been performed to find common genetic variants associated with obesity; over 250 loci have been identified where clusters of linked SNPs predispose individuals to an increase in BMI. However, little progress has been made in outlining the causal SNPs and understanding the mechanisms by which they affect the phenotype. Most of these SNPs are found in non-coding regions of the genome, and maps of relevant functional elements potentially affected by these SNPs are required. In particular,
because obesity-associated loci appear to be most often located near genes expressed in the CNS, the capacity to map genomic regulatory regions in subpopulations of neurons implicated in the control of energy homeostasis would be a first step towards linking obesity-associated SNPs to their function.

In the current study, using LepRb-expressing neurons as an example, we developed a unique protocol that can identify gene regulatory elements from a relevant subpopulation of neurons, thus providing a resource to start addressing the functional consequences of human obesity-associated SNPs. First, we demonstrate that FACS of a genetically labelled subpopulation of hypothalamic neurons results in a highly enriched nuclei population. This method is dependent on the expression of a nucleus-anchored GFP in the relevant population of neurons; therefore, it could be used for any neuronal population for which a Cre reporter is available.

The characterized LepRb-expressing neuron population is still heterogeneous and is from both sexes (it does not account for sex-specific differences). More specific subpopulations of Cre-expressing neurons could be isolated by nuclei labelling through stereotaxic injection of a Cre-dependent (double-floxed inverted orientation) adeno-associated virus expressing SUN1 superfluous GFP (sfGFP). For example, this method could be used to find regulatory regions in LepRb-expressing neurons that are more specific to the ARC nucleus or the DMH nucleus. This approach also provides the opportunity for labelling the nuclei that are expressing Cre only at the time of injection. Finally, if a Cre reporter is not available, stereotaxic injection of adeno-associated viruses expressing Cre recombinase under the control of a synapsin promoter, which confers gene expression in neurons, can also be used to label anatomical subsets of neuronal nuclei in R26-CAG-LSL-Sun1-sfGFP-myc mice.

We also developed a unique protocol that provides the ability to carry out ChIP-seq for H3K27ac on these sorted nuclei. We demonstrate the fidelity of this map to transcriptional activity and its functional relevance. However, we did not identify many leptin-treatment-enriched peaks. This result was probably due to technical and statistical constraints, that is, using an extremely low number of nuclei in these assays (Supplementary Table 1) may have led to our inability to generate statistically significant comparisons. Further development of this protocol could allow increased sensitivity and enable generalization to the study of additional histone marks or more specific transcription factors.

Demonstrating that our map is useful for linking genotype with an obesity phenotype, we highlight the overlap between our LepRb-expressing neurons’ H3K27ac peaks and GWAS SNPs at a subset of obesity-associated loci, including some with strong BMI associations. Our characterized regulatory elements provide candidate sequences for further functional analyses of these obesity-associated loci. Of note and as expected, several obesity GWAS peaks did not overlap our LepRb-expressing neuron regulatory elements, underlining the necessity to establish regulatory maps of other hypothalamic neuronal populations implicated in body weight regulation, such as melanocortin receptor 4–expressing neurons. The two-pronged
approach we describe to demonstrate the functional relevance of identified Socs3 enhancers can also be further adapted to the study of the functional effects of candidate obesity-associated SNPs on regulatory function. Finally, we also provide evidence that natural transcriptional programs that can be obtained following nuclei sorting are highly specific, can be used to study the effect of physiological changes on transcriptional activity in specific neuronal populations and, unlike TRAP-seq, allow for the study of untranslated RNA.

**Methods**

**Mice.** Lepr<sup>−/−</sup> (Lepr<sup>−/−</sup>Cre<sup>1</sup>) heterozygous mice were crossed with SUN1<sup>−/−</sup> GFP (Gt(Rosa)26Sor<sup>2</sup> <sup>1</sup> Cre<sup>1</sup> <sup>1</sup> <sup>1</sup>9neo<sup>Cre<sup>1</sup></sup>/LoxP) homozygous mice, both on a mixed 129S1/ Sv and C57BL/6J background, to generate Lepr<sup>−/−</sup>Cre<sup>1</sup>/GFP<sup>−/−</sup> postnatal day 1 (P1) heterozygous embryos. Heterozygous embryos were genotyped by PCR using the primer sequences described in Supplementary Table 4. Mice were fed Picobar mouse diet 20 5058 ad libitum for the whole study; this contained 20% protein, 9% fat and 4% fibre. Calories were provided by: protein 23.210%; fat (ether extract) 21.595%; and carbohydrates 55.231%. Lepr<sup>−/−</sup>Cre<sup>1</sup>/GFP<sup>−/−</sup> mice between 2 and 4 months old were then randomly chosen (see Supplementary Table 1) for the following treatments: (1) mice were fed ad libitum for 36 h; (2) mice were fasted for 36 h and were injected with SSS three times (12, 24 and 34 h after the beginning of fasting); (3) mice were fasted for 36 h and were injected with LLL (5 mg kg<sup>−1</sup> body weight) three times (12, 24 and 34 h after the beginning of fasting); (4) four mice were fasted for 36 h and were injected with LLL (5 mg kg<sup>−1</sup> body weight) (Los Angeles Biomedical Research Institute) at 34 h. Mice were sacrificed, and the hypothalami were dissected. All hypothalami used in the following experiments were confirmed for the following conditions. Then, we further divided those genes into genes more highly expressed in GFP<sup>+</sup> cells or GFP<sup>−</sup> cells. These differentially expressed genes were then compared to those in the TRAP-seq results in Lepr<sup>−/−</sup> or Lepr<sup>−/−</sup> cells. Additionally, we carried out gene ontology and pathway enrichment analysis on differentially expressed genes using DAVID<sup>14</sup> through the RDAVIDWebService package<sup>25</sup> in Biocoord. For unsupervised clustering, we used the WGCNA<sup>22</sup> package. Genes were first filtered to remove those with <90% of the samples containing a normalized read count depth of 20 (11,125 genes were kept). Module assignment was then inspected after the blockwiseModules function was carried out with a parameter sweep for options power (values: 6, 7, 8 and 9). As with the differentially expressed genes, enrichment for gene ontology and pathways in each module’s gene set was determined with DAVID<sup>14</sup> and the RDAVIDWebService package<sup>25</sup>.

**ChiP-seq.** Twenty hypothalami were used for each individual ChiP-seq experiment and two biological replicates were analysed for each condition. Cell nuclei were isolated and pelleted as described earlier. The cell nuclei pellet was resuspended in 2 μl FACS buffer (0.15 mM spermine, 0.5 mM spermidine, cOmplete protease inhibitor and 20 mM Na butyrate in PBS). FACS was performed with a FACSVantage SE II system equipped with a 100 μm Na butyrate GFP<sup>−/−</sup> and GFP<sup>+</sup> nuclei were sorted into PBS. After FACS, the FACS buffer was added for a final volume of 10 μl; 50 μl of 1 M CaCl<sub>2</sub> and 30 μl of 1 M MgAc2 were added and incubated on ice for 5 min, and this was followed by centrifugation at 1,800g for 15 min at 4°C. The supernatant was removed and the FACS buffer supplemented with 1% formaldehyde was gently added onto the pellet to cross-link chromatin. The cross-linked pellet was then washed with FACS buffer supplemented with 125 mM glycerine without disturbing the pellet. The cross-linked chromatin was lysed in 130 μl of Buffer B (LowCell® ChiP Kit; Diagenode) supplemented with cOmplete protease inhibitor and 20 mM Na butyrate. The lysed chromatin was sheared using a Covaris S2 sonicator to obtain on average 250 base pair (bp) size fragments; 870 μl of Buffer A (LowCell® ChiP Kit) supplemented with cOmplete protease inhibitor and 20 mM Na butyrate was added, and 20 μl of the chromatin solution was saved as an input control. A mixture of 40 μl of Dynabeads Protein A (catalogue no. 10002D; Thermo Fisher Scientific) and 40 μl of Dynabeads Protein G (catalogue no. 10004D; Thermo Fisher Scientific) was washed twice with Buffer A (LowCell® ChIP<sup>−/−</sup> kit) and resuspended in 800 μl of Buffer A; 10 μg of anti-histone H3 (acetyl K27) antibody (ChIP Grade (catalogue no. ab4729; Abcam) was added to the beads and gently agitation at 4°C for 2 h. The bead–antibody complex was precipitated with a magnet and the supernatant was removed; 800 μl of shared chromatin was added to the bead–antibody complex and rotated at 4°C overnight with agitation. The immunoprecipitated chromatin was then washed three times and Buffer C once, and eluted in 100 μl of iPure elution buffer (iPure kit; v2; Diagenode). In addition, 80 μl of iPure elution buffer was added to 20 μl input DNA. DNA was purified using the iPure kit v2 according to the manufacturer’s protocol. Sequencing libraries were generated with the Accel-NGS 2S Plus DNA Library Kit (Swift Biosciences). DNA was quantified with the Qubit DNA HS Assay Kit and 2100 Bioanalyzer (Agilent Technologies) using the High-Sensitivity DNA Kit (Agilent Technologies). Massively parallel sequencing was performed on a HiSeq 4000 system. Sequencing reads were mapped to the genome by Bowtie<sup>4</sup>, allowing one mismatch per read alignment and only uniquely aligned reads (>1 - 1 m). Peaks were called against input by using MACS2 (ref. 9). Read counts per peak were identified by ENCODE<sup>26</sup> and iPSN<sup>26</sup> using the discovery rate (DR) pipeline, which establishes a P value cut-off to accept peak calls for each condition. For differential peak intensity analysis, peaks across conditions were prioritized using BEDTools<sup>27</sup>; read coverage was obtained with HTSeq<sup>31</sup>. Peaks differentially enriched for H3K27ac histone marks were then identified using DESeq2 (ref. 12). Differentially enriched peaks were tested for enrichment for novel motifs and known transcription-factor-binding sites with MEME-ChIP<sup>32</sup>; test peak regions were split into peaks either showing increased enrichment in GFP<sup>+</sup> cells or GFP<sup>−</sup> cells. The peaks were then divided into 500-bp windows or smaller using the BEDTools<sup>27</sup> markewindows command. The FASTA sequences for these regions were then extracted using BEDTools getfasta with the repeat mask selected version. The Ensembl gene model database, the MEME-ChIP<sup>32</sup> database set of windowed peaks using the following parameters: db = tools/meme 5.0.0/db/motif_databases/JASPAR/JASPAR2018_COREvertebrates_non-relevant. mème; mème-mme-mme-22–meme-maxxx:22–meme-mmffs:50;–meme-minnss:50;–meme-maxnss:500;–meme-p:12. All ChiP-seq data were deposited in the NCBI as BioProject PRJNA418998.
ATAC-seq. Five hypothalami were used for the ATAC-seq, with two biological replicates for each condition. Cell nuclei were isolated and pelleted as described earlier. The cell nuclei pellet was resuspended in 250 μl wash buffer (0.15 mM NaCl, 0.5 mM spermidine in PBS). FACS was performed with the FACSCanto II system equipped with a 78-μm nozzle. GFP- and GFP– nuclei were sorted into PBS. After FACS, 49 μl of sorted nuclei (11,200–18,000 nuclei) were mixed with 50 μl Tagment DNA buffer (Nextera DNA Sample Preparation Kit; Illumina) and 1 μl of Tagment DNA enzyme (Nextera DNA Sample Preparation Kit), and this was followed by incubation at 37 °C for 30 min. Tagmented DNA was purified with the MinElute Reaction Cleanup Kit (Qiagen). The DNA was size-selected by using SPRISelect (Beckman Coulter) according to the manufacturer’s double-size-selection protocol. The DNA:SPRISelect ratio was 5:3 for the right-side and 2:3 for the left-side selection. Library amplification was performed as described previously. The amplified library was further purified with SPRISelect as described earlier. DNA was quantified with the Qubit DNA HS assay kit and 2100 Bioanalyzer using the High Sensitivity DNA Kit. Massively parallel sequencing was performed on a HiSeq 4000 system. Sequencing reads were mapped to the genome by using Bowtie 2 (ref. 2) (options:--no-unal; -X 2008; --no discard; -no mixed; local; very-sensitive-local). After duplicate read removal, samples were merged into one with GATK and then processed with MACS2 (ref. 5); reliable peaks were identified with the ENCODE irreproducible discovery rate (EDR) pipeline. As with ChIP-seq, peaks were then partitioned with BEDTools, red coverage was obtained with HTSeq, and differentially enriched peaks were identified with DESeq2 (ref. 9). Differentially enriched ATAC-seq peaks were tested for enrichment using the GeneTrail tool and known transcription-factor-binding MEME-ChIP in a similar fashion as the ChIP-seq peaks. However, due to peak regions being generally smaller, peaks were divided into 200 bp windows or smaller using the BEDTools makewindows command. All ATAC-seq data were deposited in the NCBI as Bioproject PRJNA418098.

GWAS and GTEX analyses. Obesity-associated variants were obtained from Ghosh and Bouchard. Variants in high linkage disequilibrium (r² of at least 0.8) with obesity-associated SNPs were then obtained for the 1000 Genomes data in all five available super populations (Ad Mixed American, African, European, East Asian, South Asian) with PLINK. All lead and linkage variants from all super populations were then concatenated into a single list with their locations in the human hg19 assembly. The coordinates of these variants were then intersected with the ChIP-seq and ATAC-seq peaks after using the UCSC liftOver tool using BEDTools. We used a random permutation test to determine the statistical significance of the rate of intersection. To do this, we randomly shuffled lift-over mouse peaks by using BEDTools, after excluding regions on the ENCODE blacklist, a total of 10,000 times. For each shuffle, we recorded the number of peaks that contained an obesity-associated lead or linked variant. We then compared the observed occurrences in the ChIP-seq or ATAC-seq data to the random permutations. The raw P value was determined to be equal to the number of permutations greater than the observed +1, divided by the total number of permutations +1. Raw P values were adjusted for multiple testing using the FDR and the R p.adjust function. Hypothalamus expression quantitative trait loci variants were downloaded from the GTEx portal (https://gtexportal.org/home/datasets). Similarly to the obesity-associated variants identified through the GWAS studies, we intersected these variants on the human genome with the lifted-over ChIP-seq and ATAC-seq peak regions and carried out a random permutation test with a total of 1,000 permutations. (Fewer permutations were carried out due to the lower degree of significance identified.) Raw P values and FDR-adjusted P values were obtained as described earlier.

Luciferase assays. The three genomic regions around Sox3 (Sox3-1, Sox3-2 and Sox3-3) were amplified via PCR from the mouse genomic using specific primers (Supplementary Table 4), and cloned into the pGL4.23[luc2/minP] enhancer assay vector (Promega) using the In-Fusion HD Cloning System (catalogue no. 639647; Takara) following the manufacturer’s protocol. For the luciferase assays, mHypoA-POMC/GFP-1 cells were seeded in a 96-well plate with 1 x 10⁴ cells per well 24 h before transfection. 120 ng of plasmid per well was transfected along with 30 ng of pGL4.74 Renilla luciferase reporter vector (Promega), to correct for transfection efficiency, using the X-tremeGene HP DNA Transfection Reagent (Roche). The DNA:x-tremeGene ratio was 1:2. Firefly and Renilla luciferase activities were measured 24 h after transfection with the Dual-Luciferase Reporter Assay System (Promega) on a Synergy 2 Multi-Mode Microplate Reader (BioTek) following the manufacturer’s protocol.

CrisprCRISPR. sgRNA sequences (Supplementary Table 4) were cloned into the pGL1 plasmid (gift from J. Weissman). Intact pGL1 plasmid that contained the sgRNA sequence against the eGFP gene was also used as a negative control. sgRNA with the pGL1 vectors along with the pH-STFV-KRAB-Cas9-P2A-mCherry plasmid (catalogue no. 60954; Addgene) were used by transfecting the Lenti-Pac HIV Expression Pack Kit (GeneCopoeia). The lentivirus was packaged with the Lentiviral transduction kit (Invitrogen). Gene expression was examined by reverse transcription quantitative PCR (RT-qPCR) using the SsoFast EvaGreen Supermix (Bio-Rad Laboratories) and was carried out on a Mastercycler Realplex 2 (Eppendorf). The primers used for the qPCR are shown in Supplementary Table 4.

References.

1. Must, A. et al. The disease burden associated with overweight and obesity. JAMA 282, 1523–1529 (1999).
2. Berrington de Gonzalez, A. et al. Body-mass index and mortality among 1.46 million white adults. N. Engl. J. Med. 363, 2211–2219 (2010).
3. Ogden, C. L., Carroll, M. D., Fryar, C. D. & Flegal, K. M. Prevalence of obesity among adults and youth: United States, 2011–2014. NCHS Data Brief 219, 1–8 (2015).
4. Swinburn, B. A. et al. The global obesity pandemic: shaped by global drivers and local environments. Lancet 378, 804–811 (2014).
5. Stunkard, A. J., Foch, T. T. & Hruby, Z. A twin study of human obesity. JAMA 256, 51–54 (1986).
6. Stunkard, A. J. et al. An adoption study of human obesity. New Engl. J. Med. 314, 193–198 (1986).
7. Silventoinen, K., Rohholm, B., Kaprio, J. & Sørensen, T. I. The genetic and environmental influences on childhood obesity: a systematic review of twin and adoption studies. Int. J. Obesity (Lond.) 34, 29–40 (2010).
8. Albuquerque, D., Stice, E., Rodriguez-Lopez, R., Manco, L. & Nobrega, C. Current review of genetics of human obesity: from molecular mechanisms to an evolutionary perspective. Mol. Genet. Genomics 290, 1191–1215 (2015).
9. Münzberg, H. & Myers, M. G. Jr. Molecular and anatomical determinants of central leptin resistance. Nat. Neurosci. 8, 566–570 (2005).
10. Turcotte, V. et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. Nat. Genet. 50, 26–41 (2018).
11. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197–206 (2015).
12. Ghosh, S. & Bouchard, C. Convergence between biological, behavioural and genetic determinants of obesity. Nat. Rev. Genet. 15, 731–748 (2017).
13. Dermitzakis, E. T. From gene expression to disease risk. Nat. Genet. 40, 492–493 (2008).
14. Schaub, M. A., Boyle, A. P., Kündaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. Genome Res. 22, 1748–1757 (2012).
15. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190–1195 (2012).
16. Heintzman, N. D. et al. Histone modifications at regulatory DNA. Nature 459, 108–112 (2009).
17. Chen, R., Wu, X., Jiang, L. & Zhang, Y. Single-cell RNA-seq reveals hypothalamic cell diversity. Cell Rep. 18, 3227–3241 (2017).
18. Leshan, L. R., Björnholm, M., Münzberg, H. & Myers, M. G. Jr. Leptin receptor signaling and action in the central nervous system. Obesity (Silver Spring) 14, 2085–2125 (2006).
19. Mo, A. et al. Epigenomic signatures of neuronal deafness in the mammalian brain. Nature 536, 1369–1374 (2015).
20. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008).
21. Cady, G. et al. Hypothalamic growth hormone receptor (GHr) controls hepatic glucose production in nutrient-sensing leptin receptor (LepRb) expressing neurons. Mol. Metab. 6, 493–405 (2017).
22. Allison, M. B. et al. TRAP-seq defines markers for novel populations of hypothalamic and brainstem LepRb neurons. Mol. Metab. 4, 299–305 (2015).
23. Allison, M. B. et al. Defining the transcriptional targets of leptin reveals a role for AIf3 in leptin action. Diabetes 67, 1093–1104 (2018).
24. Howard, J. & Flier, J. S. Attenuation of leptin and insulin signaling by Drosophila. Diabetologia 33, 1748–1757 (2012).
25. Schick, N., Oakeley, E. J., Hynes, N. E. & Badache, A. TEL/ETV6 is a signal transducer and activator of transcription 3 (Stat3)–induced repressor of Stat3 activity. J. Biol. Chem. 279, 38778–38796 (2004).
26. Valentino, M. A. et al. A uroguanylin-GUCY2C endocrine axis regulates feeding in mice. J. Clin. Invest. 121, 3578–3588 (2011).
27. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495–501 (2010).
28. Machani, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics 27, 1696–1697 (2011).
29. Nazarians-Armavil, A., Chalmers, J. A., Lee, C. B., Ye, W. & Belsham, D. D. Cellular insulin resistance disrupts hypothalamic mHypoA-POMC/GFP neuronal signaling pathways. J. Endocrinol. 220, 13–24 (2014).
30. Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. Cell 159, 647–661 (2014).
31. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015).
32. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
33. Olza, J. et al. Leptin receptor gene variant rs11804091 is associated with BMI effects on gene expression across human tissues. Nature 550, 204–213 (2017).
34. Lonsdale, J. et al. The Genotype-Tissue Expression (GTex) project. Nat. Genet. 45, 580–585 (2013).
35. Kublouai, B. M., Gemelli, T., Tolson, K. P., Wang, Y. & Zinn, A. R. Oxytocin deficiency mediates hyperphagic obesity of Sim1 haplinsufficient mice. Mol. Endocrinol. 22, 1723–1734 (2008).
36. Kublouai, B. M., Holder, J. L., Tolson, K. P., Gemelli, T. & Zinn, A. R. SIM1 overexpression partially rescues agouti yellow and diet-induced obesity by normalizing food intake. Endocrinology 147, 4542–4549 (2006).
37. Beuckmann, C. T. et al. Expression of a poly-glutamine-ataxin-3 transgene in orexin neurons induces narcolepsy-cataplexy in the rat. J. Neurosci. 24, 4469–4477 (2004).
38. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
39. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
40. Dennis, G. Jr. et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4, P3 (2003).
41. Fresn, C. & Fernández, E. A. RDAVIDWebService: a versatile R interface to DAVID. Bioinformatics 29, 2810–2811 (2013).
42. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).
43. Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). Genome Biol. 9, R137 (2008).
44. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22, 1813–1831 (2012).
45. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
46. Anders, S., Pyl, P. T. & Huber, W. HTSeq: a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
47. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218 (2013).
48. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
49. Auton, A. et al. A global reference for human genetic variation. Nature 526, 68–74 (2015).
50. Battle, A., Brown, C. D., Engelhardt, B. E. & Montgomery, S. B. Genetic effects on gene expression across human tissues. Nature 550, 204–213 (2017).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- **n/a**

- Confirmed

  - The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
  - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
  - The statistical test(s) used AND whether they are one- or two-sided
    - Only common tests should be described solely by name; describe more complex techniques in the Methods section.
  - A description of all covariates tested
  - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
  - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
  - For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
    - Give P values as exact values whenever suitable.
  - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
  - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
  - Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

**Data collection**

No software was used to collect data.

**Data analysis**

Unix/Linux command line programs:

- Bowtie v1.2
- macs2 2.1.1.20160309
- IDR v2.0.2
- STAR STAR_2.5.2b
- PLINK v1.90b5.3
- MEME v5.0.0
- bedtools v2.25.0

Programs used in are a best described by:

- R version 3.4.4 (2018-03-15)
- Platform: x86_64-apple-darwin15.6.0 (64-bit)
- Running under: macOS Sierra 10.12.6

Matrix products: default

BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/libBLAS.dylib

LAPACK: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRlapack.dylib

locale:

[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
other attached packages:
[1] parallel stats4 stats graphics grDevices utils datasets methods base

loaded via a namespace (and not attached):
[1] colorspace_1.3-2 htmlTable_1.11.2 XVector_0.18.0 base64enc_0.1-3
[5] rstudioapi_0.7 bit64_0.9-7 splines_3.4.4 geneplotter_1.56.0
[9] knitr_1.20 Formula_1.2-2 annotate_1.56.2 GO.db_3.5.0
[13] RColorBrewer_1.1-2 gtable_0.2.0
[17] assertthat_0.2.0 labeling_0.3 htmlwidgets_1.2 bit_1.1-12
[21] glue_1.2.0 GenomeInfoDbData_1.0.0 Biostrings_2.46.0
[25] gdata_2.18.0 XML_3.98-1.1 zlibbioc_1.24.0 scales_0.5.0
[29] KEGGgraph_1.38.1 RBGL_1.54.0 rhdf5_2.22.0 yaml_2.1.18
[33] memoise_1.1.0 gridExtra_2.3 rpart_4.1-13 latticeExtra_0.6-28
[37] memoise_1.1.0 gridExtra_2.3 rpart_4.1-13 latticeExtra_0.6-28
[41] stringr_1.3.0 cowplot_0.9.2 ggplot2_2.2.1 tximport_1.6.0
[45] R6_2.2.2 Hmisc_4.1-1 DBI_0.8 pillar_1.2.1
[49] bindr_0.1.1 labeling_0.3 htmlwidgets_1.2 bit_1.1-12
[53] GSBase_1.40.1 AnnotationForge_1.20.0 plyr_1.8.4 magrittr_1.5
[57] ggplot2_2.2.1 Hmisc_4.1-1 DBI_0.8 pillar_1.2.1
[61] foreign_0.8-69 withr_2.1.2 survival_2.43-1 KEGGREST_1.18.1
[65] progress_1.1.2 loo_1.1.1 grapphviz_2.22.0 digest_0.6.15 xtable_1.8-2
[69] jmsnell_0.4.3

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-seq Bioproject: PRJNA418102
ChiP-seq and ATAC-seq Bioproject: PRJNA418098

Field-specific reporting
Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

[ ] Life sciences
[ ] Behavioural & social sciences
[ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design
All studies must disclose on these points even when the disclosure is negative.

Sample size
For each assay, at least 5 mice per RNA-seq or ATAC-seq and at least 20 mice for ChiP-seq with three biological replicates for RNA-seq (total of 60 mice) and two for ATAC-seq (total of 40 mice) and ChiP-seq (total of 160 mice) were used. Mouse numbers were chosen based on preliminary experiments showing they could allow sufficient material for these genomic assays. Replicate numbers were chosen following the ENCODE guidelines for these genomic studies.

Data exclusions
One sample from the RNA-seq data set was excluded due to low transcript diversity.

Replication
Three biological replicates for RNA-seq and two for ATAC-seq and ChiP-seq.
Replication
Luciferase assays were carried out using three independent biological experiments. Two different sgRNAs targeting the same region were used for CRISPRi. All attempts at replication were successful.

Randomization
No relevant, as we needed to know the identity of each sample prior to their analyses.

Blinding
Not relevant as we needed to know the identity of each sample prior to their analyses.

Reporting for specific materials, systems and methods
We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| Antigen | Antibodies | Eukaryotic cell lines | Palaeontology | Animals and other organisms | Human research participants | Clinical data |
|---------|------------|-----------------------|---------------|----------------------------|-----------------------------|---------------|
| n/a     | x          | x                     |               | x                          |                             |               |

Antibodies

- Antibodies used: chicken anti-GFP (abcam, ab13970), 1:1000 dilution
- goat anti chicken IgY Alexa fluor 488 (Invitrogen, A11039), 1:500 dilution
- H3K27ac antibody (ab4729, Abcam), 10 ug per reaction

Validation
- The chicken anti-GFP (abcam, ab13970) was validated by Western blot and immunohistochemistry by the manufacturer (http://www.abcam.com/gfp-antibody-ab13970.html)
- The goat anti chicken IgY Alexa fluor 488 (Invitrogen, A11039) was validated by flow cytometry, immunohistochemistry and immunofluorescence by the manufacturer (https://www.thermofisher.com/antibody/product/Goat-anti-Chicken-IgY-H-LSecondary-Antibody-Polyclonal/A-11039)
- The H3K27ac antibody (ab4729, Abcam) was validated by Western blot, immunohistochemistry, immunofluorescence and ChIP (http://www.abcam.com/histone-h3-acetyl-k27-antibody-chip-grade-ab4729.html)

Eukaryotic cell lines

- Cell line source(s): mHypoA-POMC/GFP-1 (Cellutions Biosystems Inc., catalog number CLU500)
- Authentication: mHypoA-POMC/GFP-1 cells were authenticated by Immunocytochemistry and RT-qPCR by the manufacturer (https://www.cedarlanelabs.com/Products/Detail/CLU500?lob=Cellutions)
- Mycoplasma contamination: The cell lines were not tested for mycoplasma contamination.
- Commonly misidentified lines: mHypoA-POMC/GFP-1 cell line is not listed in the ICLAC database

Animals and other organisms

- Studies involving animals: ARRIVE guidelines recommended for reporting animal research

Laboratory animals
- 129S1/Sv and C57BL/6J mice between 2-4 months old were used.
- Three males and two females per replicate were used for RNA-seq.
- Three males and two females (replicate 1) or two males and three females (replicate 2) were used for ATAC-seq.
- Ten males and ten females per replicate were used for ChIP-seq.

Wild animals
- No wild animals were used in this study.

Field-collected samples
- No field-collected samples were used in this study.

Ethics oversight
- Studies approved by the UCSF Institutional Animal Care & Use Program Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.
**ChIP-seq**

**Data deposition**

- Confirm that both raw and final processed data have been deposited in a public database such as GEO.
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

**Data access links**

May remain private before publication.

- https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112125

**Files in database submission**

All ChIP-seq data was deposited in NCBI as Bioproject PRJNA418098.

**Genome browser session**

[e.g. UCSC]

http://genome.ucsc.edu/cgi-bin/hgTracks?

hgS_doOtherUser=submit&hgS_otherUserName=BatMouse&hgS_otherUserSessionName=Leptin_mm9

**Methodology**

**Replicates**

Two replicates.

**Sequencing depth**

29.6-65.3 million total reads. 12.9-34.5 uniquely aligned reads. Single read 50bp.

**Antibodies**

H3K27ac (ab4729, Abcam)

**Peak calling parameters**

For each replicate and the merge of the replicates:

macs2 callpeak -t $CHIPS -c $INPUTS -f BAM -g mm --keep-dup 1 --bw 500 -n $outname --nomodel --extsize 300 --slocal 2000 --local 20000 -p 0.001 --broad

then with peaks called above from the pseudo replicates:

idr --samples $rep1.broadPeak $rep2.broadPeak --peak-list $pooled.broadPeak --plot --idr-threshold 0.1 --input-file-type broadPeak --output-file $out.broadPeak

**Data quality**

The IDR pipeline ensures peaks are reproducible. The number of peaks at an IDR level of .1 or better ranges between 26,529 and 40,122 for the ChIP-seq. For ATAC-seq, there were 32,377 peaks and 42,419 peaks for GFP positive and GFP negative, respectively, below the IDR of .1 threshold.

**Software**

Sequencing reads were mapped to the genome using bowtie (version noted above) allowing one mismatch per readalignment and only uniquely aligned reads (-v 1 -m 1). Peaks were called against input using MACS2 and the ENCODE IDR pipelines (version and process described above).

**Flow Cytometry**

**Plots**

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

**Methodology**

**Sample preparation**

Cell nuclei were obtained from transgenic mice that express GFP in LepRb neurons in the hypothalamus.

**Instrument**

FACS Aria II (BD Biosciences) equipped with 70 um nozzle.

**Software**

FACSDiva

**Cell population abundance**

Re run FACS. Less than 0.1% GFP positives were identified in GFP negative population and vice versa.

**Gating strategy**

Gating strategy is shown in Supplementary Figure 1. Cell nuclei with fluorescence between zero to 600 were defined as GFP negative, and those with more than 6,000 were GFP positive.