Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of IRX3 and IRX5

Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of *IRX3* and *IRX5*

Authors: Débora R. Sobreira[1, *], Amelia C. Joslin[1], Qi Zhang[1, 2], Iain Williamson[3], Grace T. Hansen[1], Kathryn M. Farris[1], Noboru J. Sakabe[1], Nasa Sinnott-Armstrong[4,5], Grazyna Bozek[1], Sharon O. Jensen-Cody[6], Kyle H. Flippo[6], Carole Ober[1], Wendy A. Bickmore[3], Matthew Potthoff[6], Mengjie Chen[1, 2], Melina Claussnitzer[5, 7], Ivy Aneas[1, *], Marcelo A. Nóbrega[1, *]

Affiliations:

[1]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA.

[2]Section of Genetic Medicine, Department of Medicine, The University of Chicago, Chicago, IL 60637, USA.

[3]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road South, Edinburgh EH4 2XU, UK.

[4]Department of Genetics, Stanford University, Stanford 94305 CA, USA.

[5]Metabolism Program and Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

[6]Department of Pharmacology, University of Iowa Carver College of Medicine, Iowa City, IA 52242, USA.

[7]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02131, USA.

*Corresponding author: Débora R. Sobreira (deborarsobreira@gmail.com), Ivy Aneas (ianeas@bsd.uchicago.edu), Marcelo A. Nóbrega (nobrega@uchicago.edu).

Abstract: While coding variants often have pleiotropic effects across multiple tissues, non-coding variants are thought to mediate their phenotypic effects by specific tissue and temporal regulation of gene expression. Here, we dissected the genetic and functional architecture of a genomic region within the *FTO* gene that is strongly associated with obesity risk. We show that multiple variants on a common haplotype modify the regulatory properties of several enhancers targeting *IRX3* and *IRX5* from megabase distances. We demonstrate that these enhancers impact gene expression in multiple tissues, including adipose and brain, and impart regulatory effects during a restricted temporal window. Our data indicate that the genetic architecture of disease-associated loci may involve extensive pleiotropy, allelic heterogeneity, shared allelic effects across tissues, and temporally-restricted effects.

One Sentence Summary: The genetic architecture of loci associated with human traits implicate pleiotropy, allelic heterogeneity and temporal-specific effects.

Main Text: Although genome-wide association studies (GWAS) have contributed extensively to complex disease mapping, our understanding of the genetic architecture and molecular mechanisms underlying most disease associations remains incomplete (*1*, *2*). Recent studies suggest pervasive pleiotropy of regulatory variants modulating gene expression across multiple tissues, impacting seemingly disparate disease phenotypes (*3*, *4*). We set out to dissect the genetic architecture and phenotypic implications of a well-studied locus associated with human obesity. GWAS have identified common variants in the *FTO* gene as the strongest genetic association with obesity in humans (*5*). Much

effort has been directed towards identifying the causal variant, gene, and tissues underlying this association. The associated region is within a large topologically associated domain (TAD) of approximately 2 Mb encompassing *FTO*, *RPGRIP1L*, and the *IRXB* cluster (including *IRX3*, *IRX5*, and *IRX6*) (*6*). As a consequence of this arrangement, the obesity-associated variants could impact the regulation of any or all of these genes. In fact, most of these genes have been independently implicated in body weight management phenotypes, leading to additional controversy within the field as to which of these genes mediate the genetic association with obesity (*7–10*). In addition, while compelling evidence implicates central nervous system phenotype such as food preference and feeding behavior underlying the association with body mass index (BMI) (*5*, *11*) alternative models involving altered thermogenesis, autonomous to adipose tissue, have also been put forth as putative mechanisms (*7*, *8*). To address these discrepancies, we applied an integrated approach to mechanistically dissect the genetic and functional architecture of the obesity GWAS signal emanating from the *FTO* locus.

To ascertain the pattern of long-range genomic interactions in the locus, we generated a comprehensive chromatin interaction map in cell types relevant to obesity. We performed in situ promoter capture Hi-C (PCHi-C) in human SGBS preadipocytes and in hypothalamic arcuate-like neurons derived from human induced pluripotent stem cells (hiPSCs). PCHi-C contact maps from both cell types, and additional 4C-seq data, revealed long-range interactions between the obesity-associated locus and promoters of *IRX3* and *IRX5*, but not those of *IRX6* or *FTO/RPGRIP1L* (Fig. 1A and fig. S1A). Similar results were obtained from a enhancer capture Hi-C dataset in primary human pre-adipocytes (fig. S2A) (*12*). Because this locus is highly conserved between humans and

mice (fig. S1B), we engineered a mouse model (mm*Fto*Δ20) harboring a 20,204 bp deletion spanning the orthologous obesity-associated interval in *Fto* (fig. S1C). Using fluorescence in situ hybridization (FISH) as an orthogonal assay to PCHi-C, we interrogated the 3D organization of this region in vivo, in mouse brains from mm*Fto*Δ20 heterozygous animals. Designing fosmid-based probes for regions encompassing the *Fto/Rpgrip1l*, *Irx3*, *Irx5*, *Irx6* promoters, and the *Fto* obesity-associated locus, as well as the region directly adjacent to the 20 kb deletion (Fig. 1B, fig. S2B), we determined the pattern of interactions between the obesity-associated region and genes in the *Fto-Irxb* locus. Consistent with our PCHi-C and 4C-seq results, FISH data from WT alleles in cerebellum revealed significantly increased colocalization (≤200 nm) of the *Irx3* and *Irx5* promoters with the *Fto* obesity-associated interval in WT alleles, compared with deletion alleles and cortex cells that do not express *Irx3* and *Irx5*. We obtained similar colocalization data in lung cells, supporting our previous observations that the obesity-associated interval harbors lung enhancers (*7*). Conversely, the 20 kb deletion had no impact on the distance between the *Fto/Rpgri1L* or *Irx6* promoters and the *Fto* obesity-associated interval (Fig. 1C and D, fig. S2C). These observations support a model of chromatin compaction at this locus with the obesity-associated region physically interacting with both *IRX3* and *IRX5* in humans and mice.

We next explored the biological relevance of these observations in vivo. We genetically engineered germline null (-/-) and heterozygous (+/-) alleles for *Irx3, Irx5,* and *Irx6* in mice (fig. S3). At 20 weeks, *Irx3*-/- animals displayed a 15-20% blunting in weight gain compared to the WT control littermates, as well as a reduction of total fat mass (10-15%), activation of molecular markers of browning in white adipose tissue (WAT), and

improved glucose tolerance, mimicking phenotypes we have previously shown (fig. S4A to C and fig. S5A, D to G) (*7*). While the most striking feature of *Irx5*-/- mice is early postnatal lethality, *Irx5* heterozygous mice (*Irx5*+/-) were viable and thrived. Similar to *Irx3*-/-, *Irx5*+/- mice exhibited an anti-obesity phenotype with 15-20% weight reduction, loss of body fat mass (5%), activation of browning in WAT, and improved glucose tolerance (fig. S4D to F and fig. S5B, H to K). *Irx6* knockout (*Irx6*-/-) mice showed none of these metabolic phenotypes (fig. S4G to I and fig. S5 C, L to O). Altogether, our in vivo mouse models support our chromatin conformation data implicating *Irx3* and *Irx5*, but not *Irx6*, as potentially mediating the genetic association with obesity.

The phenotypic impact of *IRX3* and *IRX5* on adipocyte biology has been described (*7*, *8*). Specifically, a SNP (rs1421085) modulates *IRX3* and *IRX5* expression in preadipocytes and regulates an adipose thermogenesis program (*8*). These data, however, do not provide an immediate explanation for the well-described association of variants within *FTO* with eating behavior and, more specifically, eating preferences, such as increased caloric intake (*5*, *11*, *13*). Toward that end, we have previously shown that the hypothalamic expression of a dominant-negative IRX3 isoform in mice phenocopies the organismal level metabolic phenotypes seen in germline *Irx3* null mice (*7*). To interrogate for the impact of *Irx3* in molecular and physiological brain phenotypes associated with obesity, we performed transcriptomic analysis (RNA-seq) on hypothalami from *Irx3*-/- mice and WT littermates. Gene ontology (GO) enrichment analysis showed that, of the 359 up-regulated genes, at least 103 are involved in neurodevelopment and cellular processes, such as cell communication and synaptic signaling, consistent with the well-known roles of *Irx3* in brain development (fig. S6A) (*14–16*). Using the ToppGene Suite

database, we investigated GO categories for disease links and found that the top ranked diseases associated with these differentially expressed genes are obesity, diabetes, and impaired glucose tolerance (Fig. 2A, fig. S6A and B and Table S1), supporting the notion that *Irx3* expression in brain may coordinate a genetic program involved in metabolism. To examine whether *Irx3* plays a role in food intake or macronutrient preference, we subjected a cohort of adult *Irx3*-/- and WT control littermates to a series of two-bottle choice experiments in which all mice were offered the choice between water and a range of nutritive and non-nutritive tastants (*17*). We found that obesity-resistant *Irx3*-/- mice display a reduced preference for sucrose, but not lipid or protein, compared to WT animals (Fig. 2B and fig. S6C). Altered sweet preference has not been shown as a phenotype in humans harboring risk alleles in the *FTO* obesity-associated region. To test this, we obtained GWAS summary statistics from 118,950 genotyped individuals responding to a sweet preference questionnaire from 23andMe (Table S2). A GWAS of these data indicated that SNPs within the *FTO* obesity-associated region represent the second strongest association with sweet preference in humans, with the C allele of rs1421085 associated with sweet food preference over salty ($3.6X10^{-23}$, OR=1.1) (Fig. 2C). Taken together, our in vivo mouse data establish a central nervous system role of *Irx3* in the regulation of metabolism and feeding behavior analogous to phenotypes associated with allelic variants of obesity-associated SNPs within *FTO* in humans, including alterations in consummatory behavior. Previous work has described reciprocal counterregulatory mechanisms between peripheral energy expenditure and energy intake, with perturbations in diet and nutritional status inducing long-term changes in hypothalamic neurocircuit development (*18*). Future work will determine

whether the alterations in feeding behavior in *Irx3-/-* mice result from primary, autonomous dysfunction of regulatory circuits within the central nervous system, including the hypothalamus, or are secondary to peripheral effects, through the intersection of neuro-hormonal cues from adipose and other peripheral tissues.

Having uncovered a central nervous system role of *Irx3* in metabolism and feeding behavior, we next sought to characterize the regulatory potential of obesity-associated SNPs within *FTO.* To functionally classify regulatory variants in neurons and adipocytes, thought to represent tissues that participate in the genetics of obesity in humans (*19*), we used orthogonal computational and experimental approaches. For computational regulatory variant predictions, we derived multiple variant features from sequence-based methods which harness cross-species functional sequence conservation and sequence-based regulatory evidence (*20*). Experimentally, we used a Massively Parallel Reporter Assay (MPRA) to identify variants located in enhancers in hippocampal (HT22) and preadipocyte (3T3-L1) mouse cell lines. We tested all 87 common (MAF>=5%) variants in strong linkage disequilibrium (r^2>0.8) with the lead obesity GWAS associated SNP rs1558902 (*19*). We found 21 SNPs in 3T3-L1 preadipocytes and 18 SNPs in HT22 neuronal cells located in enhancers in at least three replicates tested in each cell line (Table S3). Of these, 5 SNPs displayed allelic-specific enhancer activities in preadipocyte and/or neuronal cells. Each was located in independent enhancers spread over 31 kb (Fig. 3A, fig. S7 and Table S4 and S5). Using a luciferase reporter assay, we confirmed allele-specific enhancer properties and directional effects of 4 variants in preadipocytes, 2 of which changed regulatory activity in neuronal cells as well (Fig. 3B). Of note, 3 of the 4 SNPs map within accessible

chromatin regions in human adipose and brain tissues, assayed by the Roadmap Epigenomics Consortium (Fig. 3A). In addition, we confirmed that all accessible variants score highest across multiple, orthogonal sequence-based computational metrics, including high functional conservation scores for the variant flanking 120bp regions, as evaluated with PMCA (*21*) (Table S6), sequence-based predicted functional significance scores < 0.01, as evaluated with DeepSEA (*22*) (Table S7), and all four SNPs showed remarkably consistent allele-specific chromatin accessibility with the Basset model when comparing the experimentally-derived allelic activity in pre-adipocytes and hypothalamic neurons (Table S8).All 4 SNPs are co-inherited as one common haplotype, with each allele in the obesity-risk haplotype associated with increased enhancer activity (Fig. 3C), suggesting that they may coordinately regulate target gene expression in the same direction (LDhap tool: https:// ldlink.nci.nih.gov). Our data suggest that multiple genetic variants in this locus may regulate gene expression in both adipose and neuronal tissues. This supports a model in which GWAS signals may result from a complex genetic architecture whereby allelic heterogeneity of multiple regulatory variants in distinct regulatory elements imparts shared effects across tissues, regulating the quantitative and spatial expression of multiple genes (*23*). We next determined the impact of these enhancers on gene expression. Because all four regulatory regions with allele-specific enhancer properties map within the 20 kb region that we deleted in the mouse genome (fig. S1C), we used mm*Fto*Δ20 mice to evaluate the impact, in vivo, of this deletion on the expression of neighboring genes in adipose and brain tissues. We initially assayed the expression of genes in the *Fto-Irxb* cluster during adipocyte differentiation. We isolated primary preadipocytes from mm*Fto*Δ20 and

WT mice and observed a decreased expression of *Irx3* and *Irx5* in mm*Fto*Δ20, but not of other genes in the locus (Fig. 4A). The impact of deleting these enhancers on the expression of *Irx3* and *Irx5* was restricted to preadipocytes, with no effect on expression in mature adipocytes, as previously described (*8*).

We next assayed the impact of the 20 kb deletion on gene expression in mouse hypothalamus during embryonic development (E17) as well as in adult mice (10 weeks). At E17, the 20 kb deletion leads to downregulation of *Irx3* and *Irx5*, with no impact on the expression of *Fto* and *Irx6* (Fig. 4B). Similar to adipose, this effect was restricted to embryonic development, with no alterations in *Irx3* and *Irx5* expression in adult hypothalami. To further explore the temporally-restricted expression of *Irx3* and *Irx5* in the developing hypothalamus, we assessed single-cell gene expression across windows of mouse hypothalamic development in mice (*24*), and determined that the expression of *Irx3* and *Irx5* is highest at mid-gestation and decreases steadily afterwards, being barely detectable in adult neurons (fig. S8). The expression of *Rpgrip1L* was also decreased in hypothalami of mm*Fto*Δ20 mice (Fig. 4B), raising the possibility that regulation of *Rpgrip1l* in the brain may also contribute to obesity risk, as previously suggested (*10*).

Our data suggest that variants in multiple enhancers within the *FTO* obesity-associated region regulate the expression of multiple genes in at least two major obesity-relevant tissues, adipose and brain, in mice. Next, we tested the impact of the obesity-associated region on gene expression in human hypothalamic neuronal precursors. We first assayed the dynamic expression of *IRX3* and *IRX5* during differentiation of human iPSCs into hypothalamic neurons and observed that *IRX3* and *IRX5* expression is highly

correlated and peaks at an early stage of hypothalamic neuronal differentiation, decreasing at later developmental stages, paralleling the observations in mice (fig. S9A and B). These data further support the possibility that some of the allelic effects of obesity-associated SNPs on gene expression may involve developmental phenotypes restricted to specific temporal windows and not detected in differentiated, adult tissues. A recent report uncovered evidence that the *FTO* locus variants have effects on BMI in early childhood (*25*), further raising the prospect that the association with BMI may involve a combination of developmental and growth phenotypes.

To test the effect of modulating these enhancers in a model of human hypothalamic neurons, we subsequently generated in human iPSCs a genomic deletion of a 36,100 bp segment which encompasses the *FTO* obesity-associated locus and corresponds to the deletion engineered in mm*Fto*Δ20 mice. We differentiated WT and 36,100 bp deletion (hs*FTO*Δ36) iPSCs into hypothalamic arcuate-like neurons (fig. S9C and D) (*26–28*). We performed single cell RNA sequencing (scRNA-seq) in 91,825 cells at the neuron progenitor stage to assess transcriptome differences between WT and hs*FTO*Δ36 cells. Single-cell transcriptomic profiling identified distinct cell populations within the hypothalamic neuron precursor stage, grouped into distinct subtypes. We defined different developmental stages and cell types based on the expression of known neuronal markers (*29*, *30*). Cell subtypes were designated as (1) hypothalamic neurons at late development time point (Late Dev), (2) hypothalamic neurons at an early developmental time point (Early Dev), (3) hypothalamic progenitor cells (HPC), and (4) radial glia, together all four subtypes constituting the neurogenic lineage (Fig. 4C and Table S9). We found *IRX3* and *IRX5* expressed in all hypothalamic cell subtypes. To

assay for alterations in gene expression in cellular sub-groups, we clustered cells based on the expression of 8 major neural and hypothalamic markers, including *ARNT2, NES, NEUROD1, NHLH2, NKX2-1, NPY, OTP* and *POMC* (fig. S10). We found that only in cells expressing *POMC*, which is critical in regulating normal feeding behavior and energy homeostasis, the deletion of the 36 kb resulted in reduced expression of *IRX3* and *IRX5* compared to WT cells, supporting our findings in mouse hypothalami (Fig. 4D). No other gene in the locus was differentially expressed between the two groups in any other cell type cluster. While we performed our analysis in hypothalamic cells, there currently is no clear delineation of the precise brain cell populations in which the expression of *IRX3* and *IRX5* is regulated by enhancers and allelic variants within these enhancers in the obesity-associated region. Future work tackling this outstanding question will be critical to demarcate the molecular, cellular, and organismal phenotypes involved in obesity susceptibility in this locus.

Taken together, our data highlight the complexities that arise during the functional dissection of disease-associated loci in humans. Recent work has suggested extensive pleiotropy of loci, SNPs, and gene sets underlying associations with polygenic traits in humans (*4*). Also, GTEx has shown that the regulatory effects of eQTLs tend to be highly shared across tissues (*31*). Furthermore, the impact of regulatory variants on molecular phenotypes is often dependent on developmental context, with changes in gene expression restricted to specific temporal windows (*32*). Our findings support all these observations, demonstrating how the collective effects of regulatory variants are integrated across tissues and developmental stages and result in a convergence of

phenotypes reminiscent of homeostatic mechanisms governing complex physiological traits in vivo, such as body weight regulation.

There are important limitations to our study. The choice of immortalized cell lines for the reporter assays may mask allelic effects of SNPs that would be seen in primary cells. Also, the manipulation of candidate genes in mice may result in organismal phenotypes that are quantitatively and qualitatively different than the small effect phenotypes elicited by allelic variants of SNPs associate with the human trait. Finally, the congruent macronutrient preference phenotypes we describe between *Irx3-/-* mice and humans represent but a subset of the feeding behavior phenotypes associated with this locus in humans. This may reflect species differences in the function of these genes, but also that there are other functions associated with *IRX3, IRX5* or other genes in the locus (*RPGRIP1L* or *FTO*) that contribute to the BMI association in humans.

Our work suggests that the genetic architecture of a disease-associated locus may include allelic heterogeneity, with multiple variants modifying the regulatory properties of distinct enhancers with broad tissue-specificity and regulating multiple genes in limited temporal windows. These insights provide a mechanistic framework to explain the genetic and functional architecture of GWAS loci, predicting that it will often encompass multiple phenotypic mechanisms that ultimately converge to modulate disease susceptibility.

Figure legends

Fig. 1 Regulatory architecture of obesity-associated noncoding elements within *FTO*. (**A**) PCHi-C interactions emanating from the *FTO*, *IRX3*, *IRX5*, and *IRX6* promoters in human SBGS cells (Preadipocytes) and iPSC-derived hypothalamic arcuate-

like neurons (Neurons). The yellow strip highlights the obesity-associated interval. PCHi-C interactions are presented as gray colored arcs. Red arcs highlight interactions of *IRX3* and *IRX5* promoters with obesity-associated region. (**B**) 1.1 Mb region analyzed by FISH in mouse cerebellum encompassing *Fto, Irx3, Irx5, and Irx6* genes. Fragment deleted in the mm*Fto*Δ20 mouse is indicated in blue. Fosmids used for analysis in D are indicated in red and green. (**C**) 3D-FISH with *Fto, Irx3, Irx5,* and *Irx6* probes (red) and directly distal *FTO* obesity-associated interval (green), counterstained with DAPI (blue). Bars, 5μm. (**D**) Box plots represent the distribution of interprobe distances (nm) between different probe combinations in *Irx3*-expressing (cerebellum: cer) and non-expressing (cortex: con) brain tissue of mm*Fto*Δ20 heterozygous mice. Statistical significance of differences between data sets was examined using Mann Whitney U tests. n= 50 – 60 WT and mm*Fto*Δ20 alleles each per slide. Abbreviation: FtoLD (*FTO* obesity-associated interval); FtoPr (*Fto* promoter).

Fig. 2 *Irx3* acts in the brain to regulate metabolism and changes in macronutrient selection. (**A**) Expression analysis of differentially expressed genes between hypothalami of *Irx3*-/- and WT mice using Gene Ontology (GO) annotations. Sankey flow diagram showing all genes upregulated in the hypothalami from *Irx3*-/- animals with high enrichment for Cellular Metabolic Processes and the top ranked diseases related to them. Gene symbols are shown. (**B**) Two-bottle choice experiment comparing *Irx3*-/- and WT mice. Data are expressed as mean ± SEM. *, $P < 0.05$ compared to WT group. (**C**) A regional association plot of the *FTO* locus. LocusZoom was used to plot the negative $\log_{10}$ p-value of every SNP within +/- 700 kb of rs1421085, the lead SNP in the locus.

Fig. 3 Functional variants within the *FTO* association locus modulate enhancer activity in brain and adipose. (**A**) Functional variants that showed allele-specific activity using MPRA (black boxes) and PMAC (red boxes). Colored bars indicate the chromatin state annotations from Roadmap Epigenomics Project. Tissues: adipose-derived mesenchymal stem cell cultured imputed (E025) and adipose nuclei imputed (E063); brain hippocampus middle (E071) and fetal brain male (E081). (**B**) Comparison of allele-specific activity of four variants in the *FTO* obesity-associated interval using luciferase reporter assay. The plots show the mean ± SEM from five triplicate experiments. **P* < 0.05, ***P* < 0.01, and ****P* < 0.001. (**C**) Segregation of alleles by risk or non-risk haplotype and effect on enhancer activity.

Fig. 4 Evaluation of enhancer activity in the *FTO* obesity-associated locus in neuronal and adipose tissues. Relative expression of *Rpgrip1L*, *Fto*, *Irx3*, *Ixr5*, and *Irx6* genes in (**A**) mouse preadipocyte cells, adipose tissue, and **(B)** hypothalamus. **(C)** UMAP plot showing the different cell populations identified using single cell sequencing. **(D)** Volcano plot of the differential gene expression (DE) analysis between WT and hs*FTO*Δ36 hypothalamic precursor cells with *POMC* cutoff 95 (counts) and KNN K=11. Gray dots represent genes not significantly changed. Green and red dots are genes significantly down and up-regulated, respectively. The log fold change (logFC) is shown on the x axis and the negative log10 of the adjusted *P* value is shown on the y axis (logFC cutoff > 0.6 or < -0.6, and adjusted *P* value < 0.05 as significantly differentially expressed). *IRX3* and *IRX5* are significantly differentially expressed across two conditions with KNN K ranging from 10 to 13 (*IRX3*), from 11 to 13 (*IRX5*), and cutoff value above 80 or 85 (counts). The abbreviations are: HPC (hypothalamic progenitor

cells); Early Dev (hypothalamic neurons at early development time point); Late Dev (hypothalamic neurons at late development time point). For qPCR analysis error bars represent the mean ±SEM. *$P < 0.05$ and **$P < 0.01$ compared to WT.

References and Notes

1. N. J. Timpson, C. M. T. Greenwood, N. Soranzo, D. J. Lawson, J. B. Richards, Genetic architecture: The shape of the genetic contribution to human traits and disease. *Nat. Rev. Genet.* (2018), , doi:10.1038/nrg.2017.101.
2. P. M. Visscher, N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, J. Yang, 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am. J. Hum. Genet.* (2017), , doi:10.1016/j.ajhg.2017.06.005.
3. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* (2020), doi:10.1126/science.aaz1776.
4. K. Watanabe, S. Stringer, O. Frei, M. Umićević Mirkov, C. de Leeuw, T. J. C. Polderman, S. van der Sluis, O. A. Andreassen, B. M. Neale, D. Posthuma, A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* (2019), doi:10.1038/s41588-019-0481-0.
5. J. E. Cecil, R. Tavendale, P. Watt, M. M. Hetherington, C. N. A. Palmer, An Obesity-Associated *FTO* Gene Variant and Increased Energy Intake in Children. *N. Engl. J. Med.* **359**, 2558–2566 (2008).
6. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* (2012), doi:10.1038/nature11082.
7. S. Smemo, J. J. Tena, K.-H. Kim, E. R. Gamazon, N. J. Sakabe, C. Gómez-Marín,

I. Aneas, F. L. Credidio, D. R. Sobreira, N. F. Wasserman, J. H. Lee, V. Puviindran, D. Tam, M. Shen, J. E. Son, N. A. Vakili, H.-K. Sung, S. Naranjo, R. D. Acemel, M. Manzanares, A. Nagy, N. J. Cox, C.-C. Hui, J. L. Gomez-Skarmeta, M. a Nóbrega, Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. **507**, 371–5 (2014).

8. M. Claussnitzer, S. N. Dankel, K.-H. Kim, G. Quon, W. Meuleman, C. Haugen, V. Glunk, I. S. Sousa, J. L. Beaudry, V. Puviindran, N. A. Abdennur, J. Liu, P.-A. Svensson, Y.-H. Hsu, D. J. Drucker, G. Mellgren, C.-C. Hui, H. Hauner, M. Kellis, FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.* **373**, 895–907 (2015).

9. J. Fischer, L. Koch, C. Emmerling, J. Vierkotten, T. Peters, J. C. Brüning, U. Rüther, Inactivation of the Fto gene protects from obesity. *Nature* (2009), doi:10.1038/nature07848.

10. G. Stratigopoulos, L. C. Burnett, R. Rausch, R. Gill, D. B. Penn, A. A. Skowronski, C. A. LeDuc, A. J. Lanzano, P. Zhang, D. R. Storm, D. Egli, R. L. Leibel, Hypomorphism of Fto and Rpgrip1l causes obesity in mice. *J. Clin. Invest.* (2016), doi:10.1172/JCI85526.

11. J. Wardle, C. Llewellyn, S. Sanderson, R. Plomin, The FTO gene and measured food intake in children. *Int. J. Obes.* **33**, 42–45 (2009).

12. J. G. S. Madsen, M. S. Madsen, A. Rauch, S. Traynor, E. L. Van Hauwaert, A. K. Haakonsson, B. M. Javierre, M. Hyldahl, P. Fraser, S. Mandrup, Highly interconnected enhancer communities control lineage-determining genes in human mesenchymal stem cells. *Nat. Genet.* (2020), doi:10.1038/s41588-020-0709-z.

13. L. M. Ranzenhofer, L. E. S. Mayer, H. A. Davis, H. K. Mielke-Maday, H. McInerney, R. Korn, N. Gupta, A. J. Brown, J. Schebendach, M. Tanofsky-Kraff, V. Thaker, W. K. Chung, R. L. Leibel, B. T. Walsh, M. Rosenbaum, The FTO Gene and Measured Food Intake in 5- to 10-Year-Old Children Without Obesity. *Obesity* (2019), doi:10.1002/oby.22464.

14. J. L. Gömez-Skarmeta, R. D. Del Corral, E. De La Calle-Mustienes, D. Ferrés-Marcó, J. Modolell, Araucan and Caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. *Cell* (1996), doi:10.1016/S0092-8674(00)81085-5.

15. E. J. Bellefroid, A. Kobbe, P. Gruss, T. Pieler, J. B. Gurdon, N. Papalopulu, Xiro3 encodes a xenopus homolog of the Drosophila Iroquois genes and functions in neural specification. *EMBO J.* (1998), doi:10.1093/emboj/17.1.191.

16. I. Anselme, C. Laclef, M. Lanaud, U. Rüther, S. Schneider-Maunoury, Defects in brain patterning and head morphogenesis in the mouse mutant Fused toes. *Dev. Biol.* (2007), doi:10.1016/j.ydbio.2006.12.025.

17. S. Von Holstein-Rathlou, L. D. Bondurant, L. Peltekian, M. C. Naber, T. C. Yin, K. E. Claflin, A. I. Urizar, A. N. Madsen, C. Ratner, B. Holst, K. Karstoft, A. Vandenbeuch, C. B. Anderson, M. D. Cassell, A. P. Thompson, T. P. Solomon, K. Rahmouni, S. C. Kinnamon, A. A. Pieper, M. P. Gillum, M. J. Potthoff, FGF21 mediates endocrine control of simple sugar intake and sweet taste preference by the liver. *Cell Metab.* (2016), doi:10.1016/j.cmet.2015.12.003.

18. M. C. Vogt, L. Paeger, S. Hess, S. M. Steculorum, M. Awazawa, B. Hampel, S. Neupert, H. T. Nicholls, J. Mauer, A. C. Hausen, R. Predel, P. Kloppenburg, T. L.

Horvath, J. C. Brüning, Neonatal insulin action impairs hypothalamic neurocircuit formation in response to maternal high-fat feeding. *Cell* (2014), doi:10.1016/j.cell.2014.01.008.

19. A. E. Locke, B. Kahali, S. I. Berndt, A. E. Justice, T. H. Pers, F. R. Day, C. Powell, S. Vedantam, M. L. Buchkovich, J. Yang, D. C. Croteau-Chonka, T. Esko, T. Fall, T. Ferreira, S. Gustafsson, Z. Kutalik, J. Luan, R. Mägi, J. C. Randall, T. W. Winkler, A. R. Wood, T. Workalemahu, J. D. Faul, J. A. Smith, J. H. Zhao, W. Zhao, J. Chen, R. Fehrmann, Å. K. Hedman, J. Karjalainen, E. M. Schmidt, D. Absher, N. Amin, D. Anderson, M. Beekman, J. L. Bolton, J. L. Bragg-Gresham, S. Buyske, A. Demirkan, G. Deng, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, A. Goel, J. Gong, A. U. Jackson, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, M. Mangino, I. M. Leach, C. Medina-Gomez, S. E. Medland, M. A. Nalls, C. D. Palmer, D. Pasko, S. Pechlivanis, M. J. Peters, I. Prokopenko, D. Shungin, A. Stančáková, R. J. Strawbridge, Y. J. Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang, A. Isaacs, E. Albrecht, J. Ärnlöv, G. M. Arscott, A. P. Attwood, S. Bandinelli, A. Barrett, I. N. Bas, C. Bellis, A. J. Bennett, C. Berne, R. Blagieva, M. Blüher, S. Böhringer, L. L. Bonnycastle, Y. Böttcher, H. A. Boyd, M. Bruinenberg, I. H. Caspersen, Y.-D. I. Chen, R. Clarke, E. W. Daw, A. J. M. de Craen, G. Delgado, M. Dimitriou, A. S. F. Doney, N. Eklund, K. Estrada, E. Eury, L. Folkersen, R. M. Fraser, M. E. Garcia, F. Geller, V. Giedraitis, B. Gigante, A. S. Go, A. Golay, A. H. Goodall, S. D. Gordon, M. Gorski, H.-J. Grabe, H. Grallert, T. B. Grammer, J. Gräßler, H. Grönberg, C. J.

Groves, G. Gusto, J. Haessler, P. Hall, T. Haller, G. Hallmans, C. A. Hartman, M. Hassinen, C. Hayward, N. L. Heard-Costa, Q. Helmer, C. Hengstenberg, O. Holmen, J.-J. Hottenga, A. L. James, J. M. Jeff, Å. Johansson, J. Jolley, T. Juliusdottir, L. Kinnunen, W. Koenig, M. Koskenvuo, W. Kratzer, J. Laitinen, C. Lamina, K. Leander, N. R. Lee, P. Lichtner, L. Lind, J. Lindström, K. S. Lo, S. Lobbens, R. Lorbeer, Y. Lu, F. Mach, P. K. E. Magnusson, A. Mahajan, W. L. McArdle, S. McLachlan, C. Menni, S. Merger, E. Mihailov, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, A. Mulas, G. Müller, M. Müller-Nurasyid, A. W. Musk, R. Nagaraja, M. M. Nöthen, I. M. Nolte, S. Pilz, N. W. Rayner, F. Renstrom, R. Rettig, J. S. Ried, S. Ripke, N. R. Robertson, L. M. Rose, S. Sanna, H. Scharnagl, S. Scholtens, F. R. Schumacher, W. R. Scott, T. Seufferlein, J. Shi, A. V. Smith, J. Smolonska, A. V Stanton, V. Steinthorsdottir, K. Stirrups, H. M. Stringham, J. Sundström, M. A. Swertz, A. J. Swift, A.-C. Syvänen, S.-T. Tan, B. O. Tayo, B. Thorand, G. Thorleifsson, J. P. Tyrer, H.-W. Uh, L. Vandenput, F. C. Verhulst, S. H. Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, H. R. Warren, D. Waterworth, M. N. Weedon, L. R. Wilkens, C. Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, T. L. C. LifeLines Cohort Study, E. P. Brennan, M. Choi, Z. Dastani, A. W. Drong, P. Eriksson, A. Franco-Cereceda, J. R. Gådin, A. G. Gharavi, M. E. Goddard, R. E. Handsaker, J. Huang, F. Karpe, S. Kathiresan, S. Keildson, K. Kiryluk, M. Kubo, J.-Y. Lee, L. Liang, R. P. Lifton, B. Ma, S. A. McCarroll, A. J. McKnight, J. L. Min, M. F. Moffatt, G. W. Montgomery, J. M. Murabito, G. Nicholson, D. R. Nyholt, Y. Okada, J. R. B. Perry, R. Dorajoo, E. Reinmaa, R. M.

Salem, N. Sandholm, R. A. Scott, L. Stolk, A. Takahashi, T. Tanaka, F. M. van ’t Hooft, A. A. E. Vinkhuyzen, H.-J. Westra, W. Zheng, K. T. Zondervan, T. Adipog. ADIPOGen Consortium, T. A.-B. W. AGEN-BMI Working Group, T. Cardiogram. CARDIOGRAMplusC4D Consortium, T. Ckdg. CKDGen Consortium, T. GLGC, T. ICBP, T. M. MAGIC Investigators, T. M. MuTHER Consortium, T. Mig. MIGen Consortium, T. P. PAGE Consortium, T. R. ReproGen Consortium, T. G. GENIE Consortium, T. I. E. International Endogene Consortium, A. C. Heath, D. Arveiler, S. J. L. Bakker, J. Beilby, R. N. Bergman, J. Blangero, P. Bovet, H. Campbell, M. J. Caulfield, G. Cesana, A. Chakravarti, D. I. Chasman, P. S. Chines, F. S. Collins, D. C. Crawford, L. A. Cupples, D. Cusi, J. Danesh, U. de Faire, H. M. den Ruijter, A. F. Dominiczak, R. Erbel, J. Erdmann, J. G. Eriksson, M. Farrall, S. B. Felix, E. Ferrannini, J. Ferrières, I. Ford, N. G. Forouhi, T. Forrester, O. H. Franco, R. T. Gansevoort, P. V Gejman, C. Gieger, O. Gottesman, V. Gudnason, U. Gyllensten, A. S. Hall, T. B. Harris, A. T. Hattersley, A. A. Hicks, L. A. Hindorff, A. D. Hingorani, A. Hofman, G. Homuth, G. K. Hovingh, S. E. Humphries, S. C. Hunt, E. Hyppönen, T. Illig, K. B. Jacobs, M.-R. Jarvelin, K.-H. Jöckel, B. Johansen, P. Jousilahti, J. W. Jukema, A. M. Jula, J. Kaprio, J. J. P. Kastelein, S. M. Keinanen-Kiukaanniemi, L. A. Kiemeney, P. Knekt, J. S. Kooner, C. Kooperberg, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C. Langenberg, L. Le Marchand, T. Lehtimäki, V. Lyssenko, S. Männistö, A. Marette, T. C. Matise, C. A. McKenzie, B. McKnight, F. L. Moll, A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, P. A. F. Madden, G. Pasterkamp, J. F. Peden, A. Peters, D.

S. Postma, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, P. M. Ridker, J. D. Rioux, M. D. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, H. Schunkert, P. E. H. Schwarz, P. Sever, A. R. Shuldiner, J. Sinisalo, R. P. Stolk, K. Strauch, A. Tönjes, D.-A. Trégouët, A. Tremblay, E. Tremoli, J. Virtamo, M.-C. Vohl, U. Völker, G. Waeber, G. Willemsen, J. C. Witteman, M. C. Zillikens, L. S. Adair, P. Amouyel, F. W. Asselbergs, T. L. Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, S. R. Bornstein, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C. Chambers, S. J. Chanock, R. S. Cooper, P. I. W. de Bakker, G. Dedoussis, L. Ferrucci, P. W. Franks, P. Froguel, L. C. Groop, C. A. Haiman, A. Hamsten, J. Hui, D. J. Hunter, K. Hveem, R. C. Kaplan, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W. März, M. Melbye, A. Metspalu, S. Moebus, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. A. Palmer, N. L. Pedersen, M. Perola, L. Pérusse, U. Peters, C. Power, T. Quertermous, R. Rauramaa, F. Rivadeneira, T. E. Saaristo, D. Saleheen, N. Sattar, E. E. Schadt, D. Schlessinger, P. E. Slagboom, H. Snieder, T. D. Spector, U. Thorsteinsdottir, M. Stumvoll, J. Tuomilehto, A. G. Uitterlinden, M. Uusitupa, P. van der Harst, M. Walker, H. Wallaschofski, N. J. Wareham, H. Watkins, D. R. Weir, H.-E. Wichmann, J. F. Wilson, P. Zanen, I. B. Borecki, P. Deloukas, C. S. Fox, I. M. Heid, J. R. O'Connell, D. P. Strachan, K. Stefansson, C. M. van Duijn, G. R. Abecasis, L. Franke, T. M. Frayling, M. I. McCarthy, P. M. Visscher, A. Scherag, C. J. Willer, M. Boehnke, K. L. Mohlke, C. M. Lindgren, J. S. Beckmann, I. Barroso, K. E. North, E. Ingelsson, J. N. Hirschhorn, R. J. F. Loos, E. K. Speliotes, Genetic studies of body mass index yield new insights for obesity

biology. *Nature*. **518**, 197–206 (2015).

20. N. Sinnott-Armstrong, I. S. Sousa, S. Laber, E. Rendina-Ruedy, S. E. Nitter Dankel, T. Ferreira, G. Mellgren, D. Karasik, M. Rivas, J. Pritchard, A. R. Guntur, R. D. Cox, C. M. Lindgren, H. Hauner, R. Sallari, C. J. Rosen, Y.-H. Hsu, E. S. Lander, D. P. Kiel, M. Claussnitzer, A regulatory variant at 3q21.1 confers an increased pleiotropic risk for hyperglycemia and altered bone mineral density. *Cell Metab.* (2021), doi:10.1016/j.cmet.2021.01.001.

21. M. Claussnitzer, S. N. Dankel, B. Klocke, H. Grallert, V. Glunk, T. Berulava, H. Lee, N. Oskolkov, J. Fadista, K. Ehlers, S. Wahl, C. Hoffmann, K. Qian, T. Rönn, H. Riess, M. Müller-Nurasyid, N. Bretschneider, T. Schroeder, T. Skurk, B. Horsthemke, D. Spieler, M. Klingenspor, M. Seifert, M. J. Kern, N. Mejhert, I. Dahlman, O. Hansson, S. M. Hauck, M. Blüher, P. Arner, L. Groop, T. Illig, K. Suhre, Y. H. Hsu, G. Mellgren, H. Hauner, H. Laumen, Leveraging cross-species transcription factor binding site patterns: From diabetes risk loci to disease mechanisms. *Cell* (2014), doi:10.1016/j.cell.2013.10.058.

22. J. Zhou, O. G. Troyanskaya, Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods*. **12** (2015), doi:10.1038/nmeth.3547.

23. O. Corradin, A. Saiakhova, B. Akhtar-Zaidi, L. Myeroff, J. Willis, R. Cowper-Sallari, M. Lupien, S. Markowitz, P. C. Scacheri, Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* **24** (2014), doi:10.1101/gr.164079.113.

24. R. A. Romanov, E. O. Tretiakov, M. E. Kastriti, M. Zupancic, M. Häring, S. Korchynska, K. Popadin, M. Benevento, P. Rebernik, F. Lallemend, K. Nishimori, F. Clotman, W. D. Andrews, J. G. Parnavelas, M. Farlik, C. Bock, I. Adameyko, T. Hökfelt, E. Keimpema, T. Harkany, Molecular design of hypothalamus development. *Nature* (2020), doi:10.1038/s41586-020-2266-0.

25. Ø. Helgeland, M. Vaudel, P. B. Juliusson, O. Lingaas Holmen, J. Juodakis, J. Bacelis, B. Jacobsson, H. Lindekleiv, K. Hveem, R. T. Lie, G. P. Knudsen, C. Stoltenberg, P. Magnus, J. V. Sagen, A. Molven, S. Johansson, P. R. Njølstad, Genome-wide association study reveals dynamic role of genetic variation in infant and early childhood growth. *Nat. Commun.* (2019), doi:10.1038/s41467-019-12308-0.

26. L. Yao, Y. Liu, Z. Qiu, S. Kumar, J. E. Curran, J. Blangero, Y. Chen, D. M. Lehman, Molecular Profiling of Human Induced Pluripotent Stem Cell-Derived Hypothalamic Neurones Provides Developmental Insights into Genetic Loci for Body Weight Regulation. *J. Neuroendocrinol.* (2017), doi:10.1111/jne.12455.

27. F. T. Merkle, A. Maroof, T. Wataya, Y. Sasai, L. Studer, K. Eggan, A. F. Schier, Generation of neuropeptidergic hypothalamic neurons from human pluripotent stem cells. *Dev.* (2015), doi:10.1242/dev.117978.

28. L. Wang, K. Meece, D. J. Williams, K. A. Lo, M. Zimmer, G. Heinrich, J. M. Carli, C. A. Leduc, L. Sun, L. M. Zeltser, M. Freeby, R. Goland, S. H. Tsang, S. L. Wardlaw, D. Egli, R. L. Leibel, Differentiation of hypothalamic-like neurons from human pluripotent stem cells. *J. Clin. Invest.* (2015), doi:10.1172/JCI79220.

29. J. Shin, D. A. Berg, Y. Zhu, J. Y. Shin, J. Song, M. A. Bonaguidi, G. Enikolopov,

D. W. Nauen, K. M. Christian, G. L. Ming, H. Song, Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. *Cell Stem Cell* (2015), doi:10.1016/j.stem.2015.07.013.

30. B. Artegiani, A. Lyubimova, M. Muraro, J. H. van Es, A. van Oudenaarden, H. Clevers, A Single-Cell RNA Sequencing Study Reveals Cellular and Molecular Dynamics of the Hippocampal Neurogenic Niche. *Cell Rep.* (2017), doi:10.1016/j.celrep.2017.11.050.

31. F. Aguet, A. N. Barbeira, R. Bonazzola, A. Brown, S. E. Castel, B. Jo, S. Kasela, S. Kim-Hellmuth, Y. Liang, M. Oliva, P. E. Parsana, E. Flynn, L. Fresard, E. R. Gaamzon, A. R. Hamel, Y. He, F. Hormozdiari, P. Mohammadi, M. Muñoz-Aguirre, Y. Park, A. Saha, A. V Segrć, B. J. Strober, X. Wen, V. Wucher, S. Das, D. Garrido-Martín, N. R. Gay, R. E. Handsaker, P. J. Hoffman, S. Kashin, A. Kwong, X. Li, D. MacArthur, J. M. Rouhana, M. Stephens, E. Todres, A. Viñuela, G. Wang, Y. Zou, T. Gte. Consortium, C. D. Brown, N. Cox, E. Dermitzakis, B. E. Engelhardt, G. Getz, R. Guigo, S. B. Montgomery, B. E. Stranger, H. K. Im, A. Battle, K. G. Ardlie, T. Lappalainen, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv* (2019), doi:10.1101/787903.

32. B. J. Strober, R. Elorbany, K. Rhodes, N. Krishnan, K. Tayeb, A. Battle, Y. Gilad, Dynamic genetic regulation of gene expression during cellular differentiation. *Science (80-.).* **364**, 1287–1290 (2019).

33. M. Wabitsch, R. E. Brenner, I. Melzner, M. Braun, P. Möller, E. Heinze, K. M. Debatin, H. Hauner, Characterization of a human preadipocyte cell strain with high capacity for adipose differentiation. *Int. J. Obes.* (2001),

doi:10.1038/sj.ijo.0801520.

34. P. Fischer-Posovszky, F. S. Newell, M. Wabitsch, H. E. Tornqvist, Human SGBS cells - A unique tool for studies of human fat cell biology. *Obes. Facts* (2008), , doi:10.1159/000145784.

35. Y. Li, P. Maher, D. Schubert, A role for 12-lipoxygenase in nerve cell death caused by glutathione depletion. *Neuron* (1997), doi:10.1016/S0896-6273(00)80953-8.

36. Z. Suo, M. Wu, B. A. Citron, R. E. Palazzo, B. W. Festoff, Rapid tau aggregation and delayed hippocampal neuronal death induced by persistent thrombin signaling. *J. Biol. Chem.* (2003), doi:10.1074/jbc.M301406200.

37. D. A. Knowles, C. K. Burrows, J. D. Blischak, K. M. Patterson, D. J. Serie, N. Norton, C. Ober, J. K. Pritchard, Y. Gilad, Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes. *Elife* (2018), doi:10.7554/elife.33480.

38. A. Melnikov, A. Murugan, X. Zhang, T. Tesileanu, L. Wang, P. Rogov, S. Feizi, A. Gnirke, C. G. C. Jr, J. B. Kinney, M. Kellis, E. S. Lander, T. S. Mikkelsen, Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* **30**, 271–277 (2012).

39. M. T. Maurano, E. Haugen, R. Sandstrom, J. Vierstra, A. Shafer, R. Kaul, J. A. Stamatoyannopoulos, Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet.* (2015), doi:10.1038/ng.3432.

40. J. Korhonen, P. Martinmäki, C. Pizzi, P. Rastas, E. Ukkonen, MOODS: Fast

search for position weight matrix matches in DNA sequences. *Bioinformatics* (2009), doi:10.1093/bioinformatics/btp554.

41. J. Nasser, D. T. Bergman, C. P. Fulco, P. Guckelberger, B. R. Doughty, T. A. Patwardhan, T. R. Jones, T. H. Nguyen, J. C. Ulirsch, H. M. Natri, E. M. Weeks, G. Munson, M. Kane, H. Y. Kang, A. Cui, J. P. Ray, T. M. Eisenhaure, K. Mualim, R. L. Collins, K. Dey, A. L. Price, C. B. Epstein, A. Kundaje, R. J. Xavier, M. J. Daly, H. Huang, H. K. Finucane, N. Hacohen, E. S. Lander, J. M. Engreitz, Genome-wide maps of enhancer regulation connect risk variants to disease genes. *bioRxiv* (2020), , doi:10.1101/2020.09.01.278093.

42. L. E. Montefiori, D. R. Sobreira, N. J. Sakabe, I. Aneas, A. C. Joslin, G. T. Hansen, G. Bozek, I. P. Moskowitz, E. M. McNally, M. A. Nóbrega, A promoter interaction map for cardiovascular disease genetics. *Elife* (2018), doi:10.7554/eLife.35788.

43. I. Williamson, R. Eskeland, L. A. Lettice, A. E. Hill, S. Boyle, G. R. Grimes, R. E. Hill, W. A. Bickmore, Anterior-posterior differences in HoxD chromatin topology in limb development. *Dev.* (2012), doi:10.1242/dev.081174.

44. C. Morey, N. R. Da Silva, P. Perry, W. A. Bickmore, Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development.* **134**, 909–19 (2007).

45. R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke, G. R. Abecasis, C. J. Willer, D. Frishman, in *Bioinformatics* (2011).

46. D. A. Hinds, G. McMahon, A. K. Kiefer, C. B. Do, N. Eriksson, D. M. Evans, B. St Pourcain, S. M. Ring, J. L. Mountain, U. Francke, G. Davey-Smith, N. J.

Timpson, J. Y. Tung, A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat. Genet.* (2013), doi:10.1038/ng.2686.

47. R. M. Gupta, J. Hadaya, A. Trehan, S. M. Zekavat, C. Roselli, D. Klarin, C. A. Emdin, C. R. E. Hilvering, V. Bianchi, C. Mueller, A. V. Khera, R. J. H. Ryan, J. M. Engreitz, R. Issner, N. Shoresh, C. B. Epstein, W. de Laat, J. D. Brown, R. B. Schnabel, B. E. Bernstein, S. Kathiresan, A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. *Cell* (2017), doi:10.1016/j.cell.2017.06.049.

48. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, a. D. Haussler, The Human Genome Browser at UCSC. *Genome Res.* (2002), doi:10.1101/gr.229102.

49. A. Frankish, M. Diekhans, A. M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J. M. Mudge, C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast, F. Cunningham, T. Di Domenico, S. Donaldson, I. T. Fiddes, C. García Girón, J. M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, T. Hunt, O. G. Izuogu, J. Lagarde, F. J. Martin, L. Martínez, S. Mohanan, P. Muir, F. C. P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B. M. Schmitt, E. Stapleton, M. M. Suner, I. Sycheva, B. Uszczynska-Ratajczak, J. Xu, A. Yates, D. Zerbino, Y. Zhang, B. Aken, J. S. Choudhary, M. Gerstein, R. Guigó, T. J. P. Hubbard, M. Kellis, B. Paten, A. Reymond, M. L. Tress, P. Flicek, GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* (2019), doi:10.1093/nar/gky955.

50. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013), doi:10.1093/bioinformatics/bts635.

51. F. Wagner, Y. Yan, I. Yanai, K-nearest neighbor smoothing for high-throughput single-cell RNA-Seq data. *bioRxiv* (2017), doi:10.1101/217737.

52. A. Lun, Overcoming systematic errors caused by log-transformation of normalized single-cell RNA sequencing data. *bioRxiv* (2018), doi:10.1101/404962.

53. M. D. Luecken, F. J. Theis, Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* (2019), doi:10.15252/msb.20188746.

54. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2009), doi:10.1093/bioinformatics/btp616.

55. C. Soneson, M. D. Robinson, Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* (2018), doi:10.1038/nmeth.4612.

Acknowledgements: We thank Romain Barrès, Michelle C. Ward and Xiaochang Zhang for comments and valuable suggestions. We thank Juan Tena and Jose Luis Gómez-Skarmeta for help with the 4C-seq analyses; and Sebastian Pott for assistance with the single cell RNA-seq analysis; We would also like to thank Martin Wabitsch for his generous gift of the SGBS human preadipocyte cells and David Schubert for Murine HT22 hippocampal neuronal cell line. We thank the customers of 23andMe. Inc. for answering surveys and participating in this research. **Funding:** This research was supported by a Novo Nordisk Foundation Challenge Grant (NNF18OC0033754) to

M.A.N.; grants from the National Institutes of Health, R01HL128075, R01119577 and R01DK114661 to M.A.N. IW and WAB are supported by MRC University Unit programme grant MC_UU_00007/2. **Author contributions:** Conceptualization and Supervision M.A.N. D.R.S and I.A; Designed Experiments: I.A, and D.R.S; Massively Report Assay: A.J; Human hypothalamic single cell Analysis: Q.Z. and M.C.; FISH experiment performed and analyzed: I.W. and W.A.B.; Hi-C analysis: G.T.H.; RNA-seq analysis: N.J.S.; Contributed reagent: C.O.; Methodology: K.M.F. and G.B.; Two bottle experiment and mouse single cell RNA-seq analysis: S.O. J-C, K. H. F. and M. P.; DeepSEA, PMCA and Basset analyses: NAS and MC; wrote the manuscript with comments from all authors: M.A.N, I.A and D.R.S. **Competing interests:** The authors declare no competing interests; **Data and materials availability:** RNA-seq (E-MTAB-10186), scRNA-seq (E-MTAB-10201), PCHi-C sequencing (E-MTAB-10200), ATAC-seq (E-MTAB-10257) and 4C-seq data (E-MTAB-10195) deposited at https://www.ebi.ac.uk/arrayexpress/.

Supplementary Materials

Materials and Methods

Figs. S1 to S10

Table S5 and Tables S9 to S12

Captions for Tables S1 to S4, Tables S6 to S8 and Tables S13 to S17

Other Supplementary Materials for this manuscript includes the following:

Tables S1 to S4 (.xlsx)

Tables S6 to S8 (.xlsx)

Tables S13 to S17 (.xlsx)

Supplementary Materials for

Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of *IRX3* and *IRX5*

Débora R. Sobreira*, Amelia C. Joslin, Qi Zhang, Iain Williamson, Grace T. Hansen, Kathryn M. Farris, Noboru J. Sakabe, Nasa Sinnott-Armstrong, Grazyna Bozek, Sharon O. Jensen-Cody, Kyle H. Flippo, Carole Ober, Wendy A. Bickmore, Matthew Potthoff, Mengjie Chen, Melina Claussnitzer, Ivy Aneas*, Marcelo A. Nóbrega*

correspondence to: deborarsobreira@gmail.com (D.R.S); ianeas@bsd.uchicago.edu (I.A); nobrega@uchicago.edu (M.A.N)

This PDF file includes:

Materials and Methods
Figs. S1 to S10
Table S5 and Tables S9 to S12
Captions for Tables S1 to S4, Tables S6 to S8 and Tables S13 to S17

Other Supplementary Materials for this manuscript includes the following:

Tables S1 to S4 (.xlsx)
Tables S6 to S8 (.xlsx)
Tables S13 to S17 (.xlsx)

Materials and Methods

Mice

All mice (C57BL/6J) originally were obtained from Charles River Laboratories, Inc. (Wilmington, Massachusetts, USA). The *Irx3*-/-, *Irx5*-/-, *Irx6*-/-, and mm*Fto*Δ20 knockout mice were generated using Crispr/Cas9 genome editing. Male mice were weaned at 4 weeks of age, and weekly records of weight were maintained throughout the study. Weaned mice were housed up to five per cage. For diet-induced obesity studies, 10-week-old male mice were subjected to 55% high-fat diet (HFD) (Harlan Teklad, Cat No. TD.93075) for 10 weeks. Body weight was measured every week from 4 to 20 weeks of age, and Dual-Energy X-ray Absorptiometry (DEXA) scans were preformed using a Lunar PIXImus II (GE Medical Systems) in order to analyze the animals body composition: body length (cm), bone mineral density (BMD, g/cm^2), body fat mass (%). Prior to scanning, animals were anesthetized with ketamine, and their body length was determined. Mice were housed on a 12-hour light/dark cycle with ad libitum access to food and water according to their assigned diet. Normal Diet (ND) groups received 14% fat, 64.8% carbohydrate, and 21.2% protein mouse chow (Harlan Teklad, Cat No. 2920X). High-fat diet (HFD) groups received 55% fat, 25.5% carbohydrate, and 16.4% protein mouse chow (Harlan Teklad, Cat No. TD.93075). All procedures were conducted with approval of the Institutional Animal Care and Use Committee (IACUC) of University of Chicago (ACUP-71656; IBC0934).

Generation of knockout animals

Irx3-/-, *Irx5*-/-, *Irx6*-/-, and mm*Fto*Δ20 knockout mice were generated using CRISPR/Cas9 genome editing. The sgRNAs used to target *Irx3*, *Irx5*, *Irx6*, and *Fto* were described in Table S10.

Genotyping

The genotypes of mutant mice were determined by PCR amplification of genomic DNA extracted from tails. PCR was performed for 30 cycles at 95 °C for 30 sec, 60 °C for 30 sec, and 72 °C for 1 min, with a final extension at 72 °C for 5 min. PCR amplification was performed using the primer sets in Table S10 and the PCR products were size-separated by electrophoresis on a 4% agarose gel for 1 h.

In vivo glucose tolerance test

Starting at 10 weeks of age, *Irx3*-/-, *Irx5*+/-, *Irx6*-/-, and WT littermates were fed a high fat diet (55% fat, Harlan Teklad) for 10 weeks. After this 10-week period, the animals were tested for glucose sensitivity by intraperitoneal glucose tolerance test (IPGTT). Prior to IPGTT, mice were fasted for 4h and an initial blood glucose reading was taken. This fast was followed by intraperitoneal injection of 2 mg/kg dextrose (Millipore Sigma, CAS 50-99-7), and subsequent blood glucose checks using an AccuChek Aviva glucometer. Blood glucose readings were taken at 15, 30, 60, and 120 min after dextrose injection and the area under curve (AUC) describing blood glucose levels during IPGTT was calculated by trapezoidal rule by using the following equation: AUC = 0.25×(fasting value) + 0.5×(half-hour value) + 0.75×(1-hour value) + 0.5×(2-hour value). After IPGTT, mice resumed high fat diet.

Real time qPCR

Total RNA was isolated from preadipocytes isolated from the inguinal white fat pad (iWAT) using the RNA extraction reagent RNeasy Mini Kit (Qiagen, Cat No. 74104). cDNA synthesis was performed using SuperScript III First-Strand Synthesis System (Thermo Fisher Scientific, Cat No. 18080-044). Real time qPCR reactions were performed by using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad Laboratories, Cat No. 1725270). Real time qPCR conditions were: 95 °C for 15 min followed by 40 cycles of 15 sec at 95 °C, 30 sec at 56 °C, and 30 sec at 72 °C. Relative gene expression was determined using *Hprt* expression as an endogenous internal control. All real time RT-qPCR primers were purchased from Bio-Rad Laboratories: *Irx3* (qMmuCED0040332), *Irx5* (qMmuCID0009561), *Irx6* (qMmuCED0003544), *Adrb3* (qMmuCED0001037), *Cidea* (qMmuCID0007140), *Prdm16* (qMmuCID0010482), and *Hprt* (qMmuCID0005679).

Two-bottle choice experiment

The two-bottle tastant experiment was performed as previously described (*17*). For the two-bottle choice studies, 10-week-old male mice (wild-type and *Irx3*-/- animals) were individually caged. Mice were given ad libitum access to one bottle with water and one bottle of a test solution for 3 days. All solutions were prepared with deionized water and served at room temperature. The test solution contained one of the following: 10% sucrose (Millipore Sigma, CAS 57-50-1), 10% casein (CAS 9000-71-9), 20% intralipid (Baxter, Cat No. 2B6061), or 10 mM sucralose (Millipore Sigma, CAS 56038-13-2). The spillage from drinking tubes was estimated daily by recording the change in fluid levels of two drinking tubes that were placed on an empty cage, where one tube contained that day's test solution and the other tube contained water.

Statistical analysis

The results for real time qPCR, in vivo glucose tolerance test, body weight and composition (body length, bone mineral density and body fat mass), luciferase assay and two-bottle choice experiments are shown as mean ± SEM. Statistical significance of differences among groups was determined by Student's paired t-test, with a two-tailed distribution.

Cell culture

SGBS (Simpson–Golabi–Behmel Syndrome) preadipocytes were kindly donated by Martin Wabitsch (Ulm University Medical Center, Ulm, Germany). SGBS cell growth was performed as previously described by Wabitsch et al. (*33*) and Fischer-Posovszky et al. (*34*). The SGBS cell line was cultured in DMEM/Ham's F12 (1:1) medium (Thermo Fisher Scientific, Cat No. 31330-038) supplemented with 10% FCS (Thermo Fisher Scientific, Cat No. 10270-106), 17 μM biotin (Millipore Sigma, Cat No. B-4639), 33 μM pantothenic acid (Millipore Sigma, Cat No. P-5155), and 1% penicillin/streptomycin (Thermo Fisher Scientific, Cat No. 15140-122). We maintained cells at 37 °C and 5% CO_2. Murine HT22 hippocampal neuronal cell line was a gift from David Schubert (The Salk Institute, California, USA) (*35*). These cells were maintained in Dulbecco's modified Eagle's medium (DMEM) (Thermo Fisher Scientific, Cat No. 11995) supplemented with 10% FBS (HyClone, Cat No. sh30070.03), as previously described

(*36*). Murine 3T3-L1 preadipocyte subline (derived from 3T3 fibroblasts of albino Swiss mouse embryo) was purchased from ATCC (3T3-L1, ATCC CL-173). The cells were cultured in growth medium composed of DMEM high glucose (ATCC, Cat No. 30-2002) supplemented with 10% bovine calf serum (ATCC, Cat No. 30-2030), and 1% penicillin/streptomycin (Thermo Fisher Scientific, Cat No. 15140-122), at 37 °C in 5% CO2. Human iPSCs (hiPSCs) (*37*) were grown in Essential 8 (E8) Medium (Thermo Fisher Scientific, Cat No. A1517001) supplemented with 1X penicillin-streptomycin (Corning, Cat No. 30002C1) on Matrigel-coated tissue culture dishes (Corning, Cat No. 354230). Cells were passaged when they were ~65% confluent using enzyme-free dissociation solution (30 mM NaCl, 0.5 mM EDTA, 1X PBS minus magnesium and calcium) and maintained in E8 Medium with 10 μM Y-27632 dihydrochloride (Abcam, Cat No. ab120129) for 24 hr. Medium was replaced daily. All cell cultures routinely tested negative for mycoplasma contamination using the Universal Mycoplasma Detection Kit (ATCC, Cat No. 30-1012K).

Hypothalamic neuron differentiation

hiPSCs were differentiated into hypothalamic arcuate-like neurons, as previously described by Wang et al. (*28*). Briefly, the sonic hedgehog (SHH) (R&D Systems, Cat. No. 1845-SH) signaling pathway was activated and transforming growth factor β (TGFβ) and bone morphogenetic protein (BMP) signaling were inhibited, followed by inhibition of NOTCH pathway, leading to neuronal differentiation, ventralization, and hypothalamic differentiation. After differentiation Brain-derived neurotrophic factor (R&D Bioscience, Cat No. 248-BD) was introduced to promote neuronal maturation of pro-opiomelanocortin (POMC) and Neuropeptide Y (NPY) neurons. Cells were collected at different time points and processed for in situ promoter capture HiC (PCHi-C), total RNA extraction, and single cell RNA-seq.

RNP-mediated CRISPR/Cas9 editing of hiPSCs

The hs*FTO*Δ36 knockout hiPSC line was generated using the ribonucleoprotein (RNP) genome editing machinery purchased from Integrated DNA Technologies (IDT). The RNA oligos (crRNA) used to target the human obesity association region in the first intron of the *FTO* gene are described in Table S10. The RNA oligos were designed using IDT software and two RNA oligos were designed in order to delete the human obesity association region in the first intron of the *FTO* gene (36,100 bp). The crRNA and tracrRNA were complexed according to manufacturer's instructions. Briefly, the crRNA and tracrRNA were mixed in equimolar concentrations (50 μM of each), then incubated at room temperature for 20 minutes to form the ribonucleoprotein. The crRNA (1) and (2) were annealed with tracrRNA in individual reactions. The two RNP complexes were nucleofected into human iPSC cells using Amaxa 2b-Nucleofector device (Bioscience Lonza) with the program A23. Cells were plated into one 22 cm^2 flask. Single cell colonies were picked and screened for the presence of homozygous deletion bands using PCR. The PCR reaction was performed for 30 cycles at 95 °C for 30 sec, 60 °C for 30 sec, and 72 °C for 1 min, with a final extension at 72 °C for 5 min. PCR amplification was performed using primer set in Table S10 and PCR products were size-separated by electrophoresis on a 1% agarose gel for 1 h.

Construction of luciferase reporter assay vectors

We evaluated the enhancer activity of all the variants identified by MPRA in the *FTO* obesity-associated locus (Table S3). We synthesized 1 kb long fragments (+/-500 bp window centered on both variants, risk and non-risk) (gBlocks Gene Fragments, IDT). Both alleles (risk and non-risk) for each variant were inserted into the multiple cloning site upstream of the minimum promoter of the pGL4.23[luc2/minP] vector (Promega) and tested by luciferase reporter assay. We verified the identity of each construct clone by DNA sequencing. The assay was run in triplicates, and all constructs were transfected into mouse hippocampus (HT22) cell line and mouse preadipocyte (3T3-L1) cells. As a negative control (scramble), we used a DNA sequence (chr9:6,161,550-6,162,024) devoid of any epigenetic marks of active chromatin.

Luciferase assay

Luciferase assays were conducted in 24-well plates. Transfection with 500 ng construct DNA using Lipofectamine LTX with Plus Reagent (Thermo Fisher Scientific, Cat No. 15338030) was performed in triplicate. HT22 (24-well plate) were transfected one day after plating at approximately 80% confluence and 3T3-L1 (24-well plate) three days after plating at approximately 70% confluence. Both cell lines were co-transfected with the firefly luciferase constructs with the *Renilla* luciferase reporter vector pGL4.73[hRluc/SV40] to normalize the transfection efficiency. Twenty-four hours after transfection, cells were washed with PBS and lysed in 1x passive lysis buffer (Promega, Cat No. E1910) on a rocking platform for 30 min at room temperature. The luciferase assays used for allele specific enhancer activities had between 4 technical replicates per construct, where different DNA preps were used and the cells were transfected, collected, and analyzed on different days.

Massively Parallel Reporter Assay (MPRA)

Of the 97 lead-SNPs identified by the GIANT consortium (*19*), one (rs1558902) resides in the *FTO* obesity-associated intronic region. Using the 1000 genomes Phase 3 database in European individuals (CEU), we identified a total of 87 biallelic SNPs in strong Linkage disequilibrium (LD) (r^2 >0.8) with the rs1558902 variant in a 100 kb interval (+/- 50 kb window centered on rs1558902) that have a minor allele frequency (MAF) > 5% to test with MPRA. The MPRA library was generated as described by Melnikov et al. (*38*) with modifications described below. Each fragment was designed as such: 5 -ACTGGCCGCTTCACTG-*enh*-GGTACCTCTAGA-barcode-AGATCGGAAGAGCGTCG-3. DNA *enh* fragments were synthesized using an Agilent array at 230 bp lengths (Agilent Technologies). Each allele of the 87 variants was synthesized with 175 base pairs of human DNA context. Each allele was assigned 18-19 unique 10 bp DNA barcodes that were generated using a randomized set of base-composition matched A, C, T, or Gs. Upon receipt from Agilent array, the fragment pool was resuspended in 100 µl of nuclease-free water and Gibson Assembly homology arms were added via PCR using a Micellula DNA emulsion and Purification Kit (EURx Molecular Biology Product, Cat No. E3600-01). Using Gibson Assembly (New England Biolabs, Cat No. E2611S), these fragments were then ligated into the pMPRA1 vector (Addgene #49349). After assembling this vector, a truncated eGFP (60 bp long) containing a minimal promoter and spacer sequence (141 bp in size) was cloned in

between the *enh* fragments and barcodes using T4 DNA ligase (New England Biolabs, Cat No. M0202S) at 16ºC for 12 hr. To preserve library complexity, the efficiency of cloning transformations was maintained at $>10^8$ cfu/μg. The resulting plasmid library was transfected into at least 10 million hippocampus (HT22) or preadipocyte (3T3-L1) cells per experimental replicate. After transfection, mRNA was extracted from total RNA using Invitrogen Dynabeads (Thermo Fisher Scientific, Cat no. 61001) and treated with Promega RQ1 DNAse (Promega, Cat No. M6101). First-strand cDNA was synthesized from 1 μg mRNA using Superscript III Reverse Transcriptase (Thermo Fisher Scientific, Cat No. 18080-044). cDNA was treated with RNAse A (Thermo Fisher Scientific, Cat No. 12091-021) and RNAse T1 (Thermo Fisher Scientific, Cat No. EN0541) for one hour and then cleaned with the Qiagen Minelute PCR purification kit (Qiagen, Cat No. 28004). MPRA sequencing libraries were generated directly from the total amount of cDNA reaction. The libraries were sequenced in indexed pools using 50-nt paired-end reads on Illumina HiSeq 4000 instruments (Illumina).

MPRA data analysis

Barcode sequences were retained if they were exact matches to synthesized barcodes. Each barcode was then counted and barcodes with low expression were removed from analysis. Enhancer activity per barcode was calculated using the following equation: log2 (output(CPM) – input(CPM)). Activity was quantile normalized and a one-tailed Mann-Whitney U Test with FDR (<5%) was used to assess enhancer significance for all alleles retaining at least 7 unique barcodes. For regions where one or both alleles were considered enhancers, we then performed a two-sided Mann-Whitney U test with FDR (<5%) to determine significance of differences between alleles. MPRA experiments were comprised of two soft biological replicates with 2 technical replicates per soft biological replicate. A replicate was considered "biological" if the input DNA library was separately cloned from the beginning from our Agilent oligonucleotides.

Phylogenetic Module Complexity Analysis (PMCA).

We used the PMCA method described in (*21*) with several modifications. Briefly, 972 position weight matrices from the Catalog of Inferred Sequences of Binding Preferences (the Catalog of Inferred Sequences of Binding Preferences (CIS-BP)) were grouped in 192 motif matrix families using TomTom, as previously described (*39*), and families were further overlapped by motif name to create a many-to-many mapping where individual TFs had multiple motifs annotated. MOODS (*40*) was used to scan a variant-flanking regions of the human reference genome (variant at mid-position) and its orthologous regions for cross-species conserved groups of transcription factor binding site motifs, so called groups of transcription factor binding site motifs, so called motif modules. A module is defined as a set of binding site motifs, whose order and distance range is conserved across species (*21*). The PMCA method counts instances of conserved motifs within conserved modules within the 120bp sequence context of a given variant. Enrichments of motifs in conserved modules are computed 10,000 permutations of orthologous sets. The PMCA method counts instances of conserved motifs within conserved modules within the 120bp sequence context of a given variant. Enrichments of motifs in conserved modules are computed 10,000 permutations of orthologous sets.

Basset convolutional neural network (CNN)

ATAC-seq IDR reproducible peaks of iPSC derived neuron differentiation along with day 0, day 3, day 6, and day 14 of differentiation of immortalized AMSCs (dbGAP, PRJNA664585) (*20*), were collated and normalized to 60bp along with 163 other cell types (*41*). A Basset model was trained with three convolutional layers (256, 128, and 128 filters; 9, 7, and 5 filter sizes; 0.1 dropout, 2 width pooling) and three fully connected hidden layers with 128 units and 0.25 dropout, using weight normalization 5, a learning rate of 0.01, and momentum of 0.995. The best validation accuracy model was used for downstream analysis (*20*).

DeepSEA

SNPs in high LD (r^2 > 0.8 in 1000G EUR) with the lead variant were formatted as a VCF file and evaluated for activity using the original DeepSEA model (http://deepsea.princeton.edu/) (*22*). Functional significance scores were extracted as described, reproduced here: "the DeepSEA functional significance score for a variant is defined as the product of the geometric mean E value for predicted chromatin effects and the geometric mean E value for evolutionary conservation features" (*22*).

RNA-sequencing

Total RNA from adult mouse hypothalamus was extracted using Trizol (Millipore Sigma, Cat No. 15596018). Quality and quantity of RNA samples were measured using Agilent 2100 Bioanalyzer instrument (Agilent). RNA-seq libraries were generated using the TruSeq RNA Library Prep Kit v2 (Illumina, Cat No. RS-122-2001) and were sequenced on Illumina HiSeq 2500 platform (Illumina). Reads were aligned to mm9 with STAR 2.5.1b [PMID: 23104886] without FilterMultimapNmax 1 and exon models from the basic set of Gencode release M1 (sjdbGTFtagExonParentTranscript). Read counts per gene were obtained by assigning reads to exons based on their alignment coordinates. We used limma 3.26.8 [PMID: 25605792] with voom transformation of read counts and quantile normalization to identify differentially expressed genes with a minimum absolute fold-difference of 1.5 and adjusted *P* value of 0.05. No correction for batch effects was applied because samples clustered in control/treatment groups as expected. The human Gene Ontology (GO) associations of GO terms [PMID: 10802651] to genes and the GO database were downloaded on January 22, 2016 from http://geneontology.org/gene-associations and terms and parent terms were assigned to Gencode mouse M1 genes. A hypergeometric test was used to calculate the statistical significance of the difference of the number of genes associated with a given GO term in a particular gene set and the universe of all genes ($P < 0.05$). *P* values were corrected with the R package p.adjust function using the fdr method. Five biological replicates were performed.

In situ promoter capture Hi-C (PCHi-C)

Five million human SGBS adipocytes and hypothalamic arcuate-like neurons from iPSCs were harvested from tissue culture dishes and resuspended in formaldehyde solution, 37% (Millipore Sigma, Cat No. 104003) to a final concentration of 1%, v/v for 10 minutes at room temperature. Glycine (Millipore Sigma, CAS 56406) was added to a final concentration of 0.2 M and cells were pelleted, snap frozen in liquid nitrogen, and

stored at −80 °C. The in situ PCHi-C libraries were performed as in Montefiore et al. (*42*).

In situ PCHi-C analysis

SGBS: paired-end reads were separately aligned to hg19 with Bowtie 2 version 2.3.4.1. Aligned reads were then paired and Hi-C interactions were called with HOMER 4.9 (PMID: 20513432) with the program findHiCInteractionsByChr.pl and parameters -res 2000 -superRes 5000.

Hypothalamic arcuate-like neurons: HiCUP (version 0.5.9) was used to map PCHi-C reads to the genome and remove technical artifacts. To improve processing speed, raw fastqs were divided into 15 segments. Mapping, artifact detection, and deduplication was performed on each segment independently. Resulting bam files were then merged and HiCUP de-duplication was re-performed on the merged sample to remove PCR artifacts. CHiCAGO (version 1.14.0) was performed to identify significant (-log (weighted P value) >5) promoter-baited interactions, using a convolution background model and P value weighting to account for distance-dependent random interactions between genomic location and technical artifacts. CHiCAGO was performed independently on each replicate and on the merged set of all two replicates from each time point to identify significant chromatin interactions specific to that time point. Significance was defined as a CHiCAGO score > 5. To identify both significant interactions and interactions which do not meet CHiCAGO's significance threshold, CHiCAGO was run once to identify only significant interactions and a second time with no threshold to report all promoter-baited interactions.

Two technical replicates for each cell line were performed derived. Each technical replicate was analyzed alone, and additionally the technical replicate raw sequencing data was merged and analyzed to produce merged datasets. Merged datasets and individual replicate datasets were used in downstream analyses.

Fluorescence in situ hybridization (FISH)

Whole mount mouse brains were fixed in 4% paraformaldehyde (PFA) overnight at 4 ºC (Millipore Sigma, Cat No. P6148). The PFA-fixed brains were permeabilized in 0.5% Triton X-100 (Millipore Sigma, Cat. No. 648463), washed in PBS, and stored at −80 °C. 3D-FISH was carried out as previously described (*43*). Fosmid clones were prepared and labelled as previously described (*44*). Between 160 and 240 ng of biotin- and digoxigenin-labeled fosmid probes were used per slide, with 16-24 μg of mouse Cot1 DNA (ThermoFisher, Cat No. 18440016) and 10 μg salmon sperm DNA. For 4-color FISH, similar quantities of the additional fosmid were labeled with Green496-dUTP (Enzo Life Sciences, Cat No. ENZ-42831). Fosmid clones used as FISH probes are listed in Table S11 and genome co-ordinates were taken from the USCS mouse genome browser (NCBI37/mm9).

Image analysis of FISH

Slides were imaged using a Photometrics Coolsnap HQ2 CCD camera and a Zeiss AxioImager A1 fluorescence microscope with a Plan Apochromat 100x 1.4NA objective, a Nikon Intensilight Mercury based light source (Nikon UK Ltd.), and either Chroma #89014ET (3 color) or #89000ET (4 color) single excitation and emission filters (Chroma

Technology Corp.) with the excitation and emission filters installed in Prior motorized filter wheels. A piezoelectrically driven objective mount (PIFOC model P-721, Physik Instrumente GmbH & Co.) was used to control movement in the z dimension. Step size for z stacks was set at 0.2 µm. Hardware control, image capture, and analysis were performed using Nikon Nis-Elements software (Nikon UK Ltd.). Images were deconvolved using a calculated point spread function with the constrained iterative algorithm of Volocity (PerkinElmer Inc.). Image analysis was carried out using the Quantitation module of Volocity (PerkinElmer Inc.).

GWAS sweet versus salty

GWAS summary statistics for sweet versus salty preference were obtained from 23andMe based on 118,950 participants' answers to the question "When you're in the mood for a snack, what kind of snack do you usually reach for?". Participants could choose one of four responses (Sweet / Salty or savory / Both / Neither), and only those that responded "Sweet" or "Salty or savory" were included. The results are visualized in a regional association plot generated by LocusZoom (*45*) centered on rs1421085, at the *FTO* locus. The negative log10 of the *P* value of all genotyped SNPs in the region was plotted on the y-axis and the color of each SNP represents the pairwise linkage disequilibrium (LD) of that SNP with rs1421085. LD estimates were generated based on the 1000 Genomes Project EUR population. The GWAS was conducted by 23andMe Personal Genome Service and the single-phenotype GWAS were run as previously showed by Hinds et al. (*46*). Research participants' answers to the questionnaire were assessed. A total of 54,901 participants (46%) answered "Sweet" and 64,049 participants (54%) answered "Salty or savory." Participants who answered "Both" or "Neither" were excluded. Twenty-nine genomic loci were identified that passed criteria for genome-wide association significance (summarized in Table S12).

4C-sequencing

4C-seq was performed as previously described (*47*). In brief, single cells from three whole mouse hypothalami were crosslinked in 2% formaldehyde-PBS-10% FCS for 30 min at room temperature. After crosslinking, cells were incubated in lysis buffer for one hour, washed in 1X PBS, and resuspended in the correct enzyme buffer. SDS (to a concentration of 0.3%), Triton X-100 (to a concentration of 2.6%), and 200 units of the first enzyme were added for a 12 hours incubation in a thermomixer at 37 °C and 900 rpm; The enzyme were heat-inactivated and first ligation was performed, adding 50 units of ligase to a volume of 7 mL and then incubating overnight at 16 °C. After first ligation, 50 units of the second enzyme were added and cells were incubated for 4 hours in a thermomixer at 37 °C and 900 rpm, followed by heat inactivation of the enzyme (65 °C for 25 min). Samples were diluted with ligation buffer to a DNA concentration of < 5 ng/µl. 100 units of ligase was then added for overnight incubation at 16 °C. DNA was purified using Qiagen PCR purification kit (Qiagen, Cat No. 28104) in 10 mM Tris pH 7.5. The primers and enzymes information are in Table S10. Primers were extended with dangling Illumina adaptor sequences. The 4C-seq analyses were done as described (*7*).

In silico analysis of deposited scRNA-seq data

Single cell RNA-seq data from Romanov et al. (*24*) were downloaded from the NCBI Gene Expression Omnibus (GEO accession #: GSE132730), loaded into the R package Seurat (v3.1), and filtered using default parameters. Gene expression in each cell type at different mouse development stages (E17, P0, P2, P10, and P23) was normalized using the NormalizeData function and the expression of each gene was scaled using the ScaleData function. Only neurons which expressed *Irx3* and *Irx5* were subclustered. *Irx3* and *Irx5* expression was determined by identifying highly variable genes and was used as an input for dimensionality reduction via principal component analysis (PCA). The identified principal components were then used as an input for clustering analysis using the FindClusters function.

5-prime single-cell RNA library construction and sequencing

Cells were processed and all steps were performed following the 10X standard manufacturer's protocols. We used the Chromium Single Cell 5′ Library & Gel Bead Kit v2 (10X Genomics, Cat No. PN-1000020). Four technical replicates were run in parallel for each sample (hypothalamic progenitor cells WT and hs*FTO*Δ36). In brief, between 8000 and 10000 cells were recovered for library preparation and sequencing. All samples and reagents were prepared and loaded into the chip. Then, the Chromium Controller was run for droplet generation. Reverse transcription was conducted in the droplets. cDNA was recovered through demulsification and bead purification. Pre-amplified cDNA was further subjected to library preparation. Libraries were sequenced on an Illumina Hiseq 4000.

Single-cell RNA-seq data processing

Pre-processing: Raw sequencing data from each sample were aligned to the human reference genome (hg38) (*48*) and the GencodeV27 gene annotation (*49*) using STARsolo v2.7.3 (*50*) with default settings for 10x scRNA-seq data. Read counts were obtained for each cell barcode and annotated gene in the reference. The appropriate cell barcode file (737K-august-2016.txt) was obtained from: https://github.com/10XGenomics/cellranger/tree/master/lib/python/cellranger/barcodes. The raw matrices were used for all downstream analyses. We applied the following pre-processing procedures on the UMI counts matrix: we first filtered out cells that expressed fewer than 3 different genes, and then filtered out genes expressed in fewer than 500 cells. There were eight replicates in total done at two different times (first two replicates and then six replicates), four replicates for each of the two conditions (hypothalamic progenitor cells WT and hs*FTO*Δ36). After the pre-processing procedures, the numbers of retaining cells and genes as well as mitochondrial gene content (using a threshold of 20%) for each replicate are included in the Table S13.

Differential Expression Analysis: Merging K Nearest Neighbors to Meta Cells: Only 12.04% of all entries in the UMI count matrix (eight replicates combined together) are non-zero values. For the non-zero entries, most have a UMI count of 1. Only 4.98% of entries contain UMI counts equal to or greater than 2. To enhance signals in the sparse UMI count matrix obtained from scRNAseq, we used K nearest neighbor (KNN) (*51*) smoothing to aggregate gene-level UMI counts and form meta-cells. This was done in each replicate individually. The KNN algorithm adopts a step-wise approach, whereby initially, each profile is only minimally smoothed (using k1 = 1). In the second step, a

larger set of nearest neighbors was identified for each cell based on those minimally smoothed profiles, and the raw data were then smoothed using these larger sets of neighbors. Additional steps using increasing ki were performed until the desired degree of smoothing is reached. We used a series of K values for ranging from 8 to 13. The choices of K were heuristic. We ruled out K values smaller than 8 because such Ks did not effectively enhance the signals. We ruled out K values larger than 13 because such Ks could introduce false positives. After forming the meta-cells for each replicate, we combined all the meta-cells from the eight replicates together, and then checked differential expression across the two conditions.

Identifying cells with high expression of *POMC*: Because cells collected in this study were not fully differentiated hypothalamic neurons, we expected that a very small fraction of cells would highly express *POMC*. We tried fairly high cutoffs, where we selected the top 7.2% to 1.6% of *POMC*-enriched-cells based on the rankings of *POMC* expression. The numbers of cells across the two conditions, and the mitochondrial content after KNN and identification of cells with high expression of *POMC* can be found in the Table S14.

DE analysis: Log transformation of normalized data can introduce spurious differential expression effects into scRNA-seq data (*52*). This effect is particularly pronounced when normalization size factor distributions differ strongly between tested groups. Therefore we focused on DE analysis tools which work with counts data. A comparison study of DE analysis suggested that edgeR is among the best performing DE analysis tools for scRNA-seq data (*53*). We thus used edgeR (*54*) to analyze the aggregated meta-cells for differential expression between WT and hs*FTO*Δ36 hypothalamic progenitor cells. The analysis pipeline used in this paper was obtained from (*55*). Using a cutoff value of -0.6 for log fold change and a cutoff of 0.05 for adjusted *P* value, we found no evidence for differential expression of *RPGRIP1L* between WT and hs*FTO*Δ36 cells, regardless of the choice of K and the cutoffs for *POMC* expression. There was no evidence for differential expression of *FTO* between the two conditions either, regardless of the parameter choices. For both *IRX3* and *IRX5*, with K = 8 and 9, they were not differentially expressed between the two conditions for all *POMC* cutoffs. However, as we increased the *POMC* cutoff with those two Ks, *IRX3* and *IRX5* were more likely to be differentially expressed between WT and hs*FTO*Δ36 cells. With K = 10, 11, 12, and 13, for low *POMC* cutoffs, *IRX3* and *IRX5* were not differentially expressed between the two conditions. As we gradually increased the *POMC* cutoff, *IRX3* and *IRX5* were differentially expressed, with their expression being reduced in hs*FTO*Δ36 compared to WT cells. The logFC and *P* values corresponding to different cutoffs and different Ks are in Table S15. We also checked the following hypothalamic and neuronal markers: *ARNT2*, *NES*, *NEUROD1*, *NHLH2*, *NKX2-1*, *NPY*, *OTP*, and focused on cells with enriched expression of these genes, but we did not detect differential expression of *IRX3* and *IRX5* between the two conditions. We listed the numbers of cells in each of the two conditions and the mitochondrial content after KNN and identification of enriched cells in Table S16. We also listed the logFC and *P* values corresponding to different cutoffs and different Ks in Table S17.

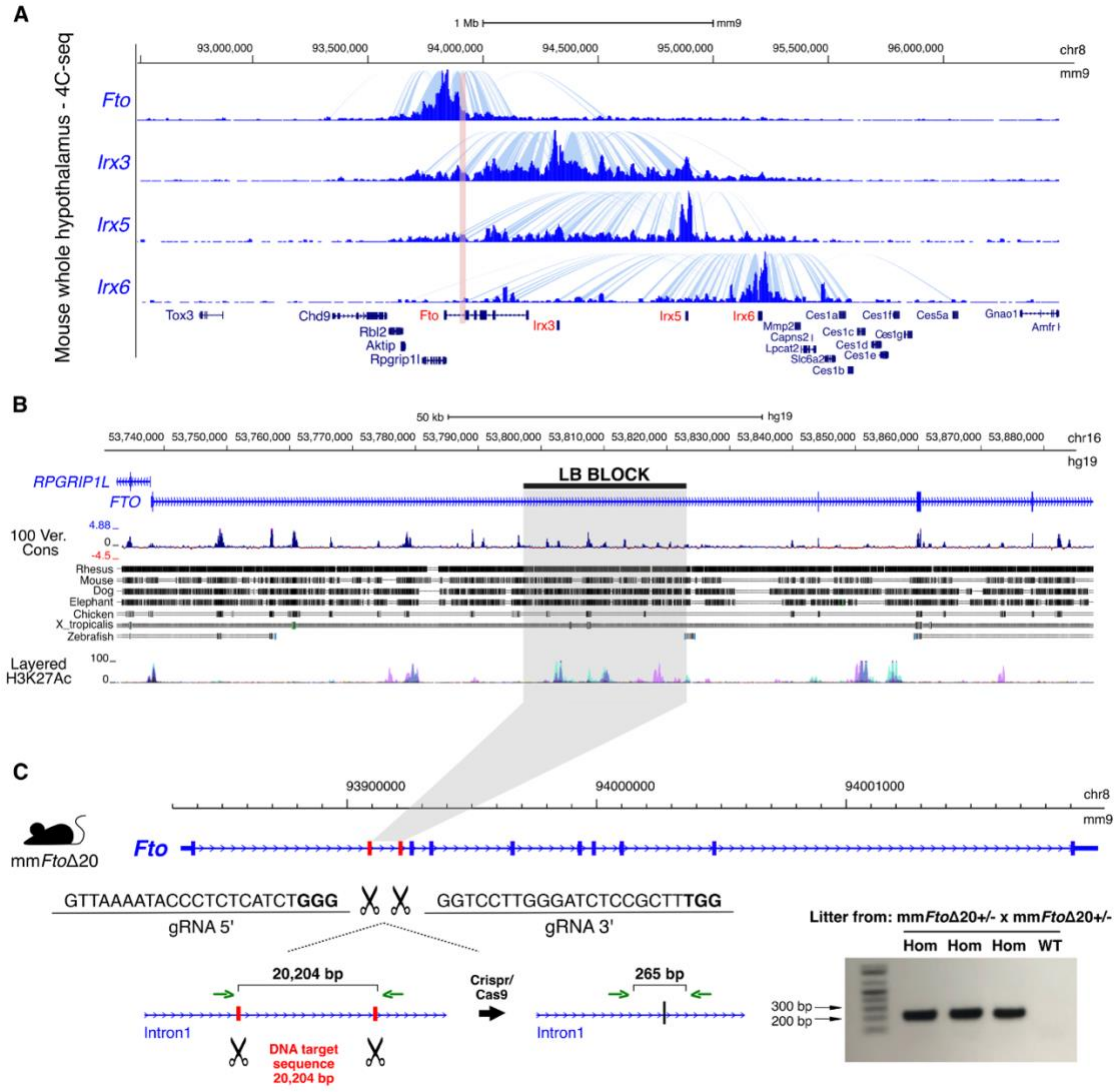

Fig. S1. Regulatory architecture of obesity-associated noncoding elements within *FTO*. (**A**) 4C-seq profiles of *Fto*, *Irx3*, *Irx5*, and *Irx6* gene promoters in adult mouse whole hypothalamus. Spider plots show statistically significant contacts (blue arcs) of each viewpoint. The obesity-associated interval is highlighted in red. (**B**) Evolutionary conservation in the obesity-associated locus within *FTO* (chr16:53731249–54975288), among different species (*7*). (**C**) Overview of the CRISPR/Cas9 strategy to delete 20,204 bp within the first intron of the *Fto* gene. The gRNA-targeting sequence (gRNA) is underlined and the PAM sequences are indicated in bold. Exons are represented as thick blue boxes, introns are indicated as blue lines with arrows, and the red boxes indicate the two DNA-targeting regions. Agarose gel shows the screening of F0 mouse tail genomic DNA. PCR analysis of the targeted locus showed a 265 bp band corresponding to the genomic deletion when the combination of both gRNAs excised the DNA target region.

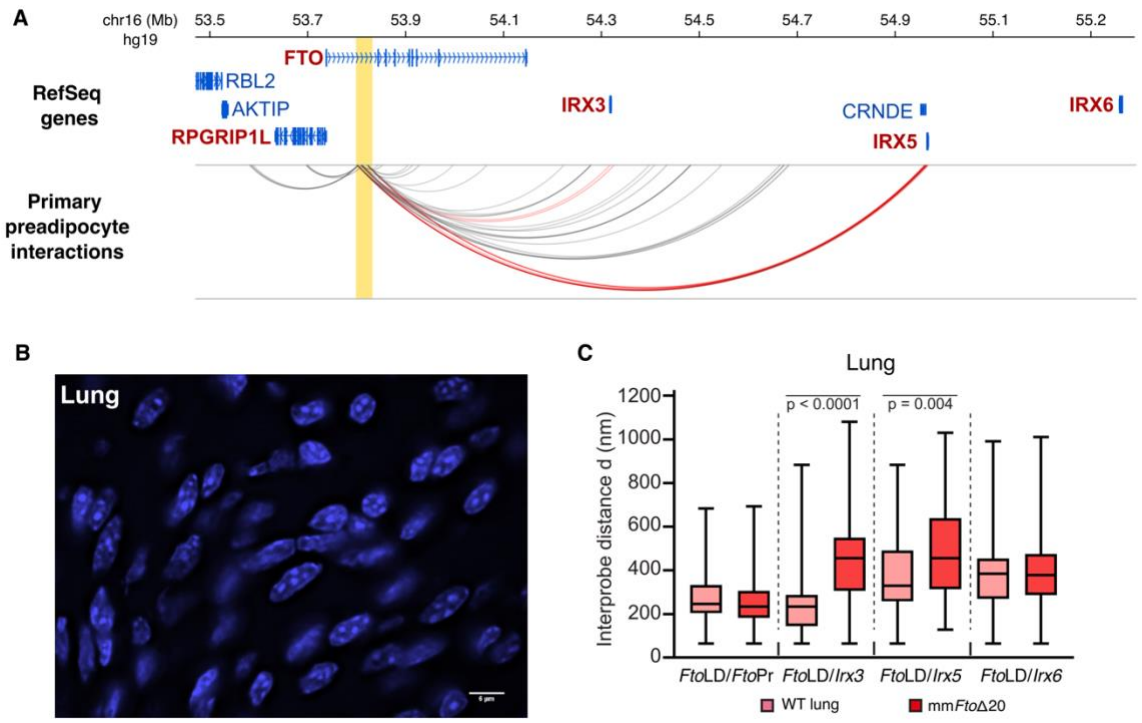

Fig. S2. Chromatin conformation of obesity-associated noncoding region within *FTO*. (**A**) Capture Hi-C interactions emanating from the obesity-associated interval within *FTO* gene in primary human preadipocytes (GEO:GSE140782) (*12*). The yellow strip highlights the obesity-associated interval. The capture Hi-C interactions are presented as gray colored arcs. Red arcs highlight interactions of obesity-associated region with *IRX3* and *IRX5* promoters. (**B**) Representative image of dapi-stained tissue sections derived from lung. Bars 6 μm. (**C**) Box plots represent the distribution of interprobe distances (nm) between different probe combinations between WT and mm*Fto*Δ20 heterozygous mice in lung tissue. n= 50 – 60 WT and mm*Fto*Δ20 alleles each per slide.

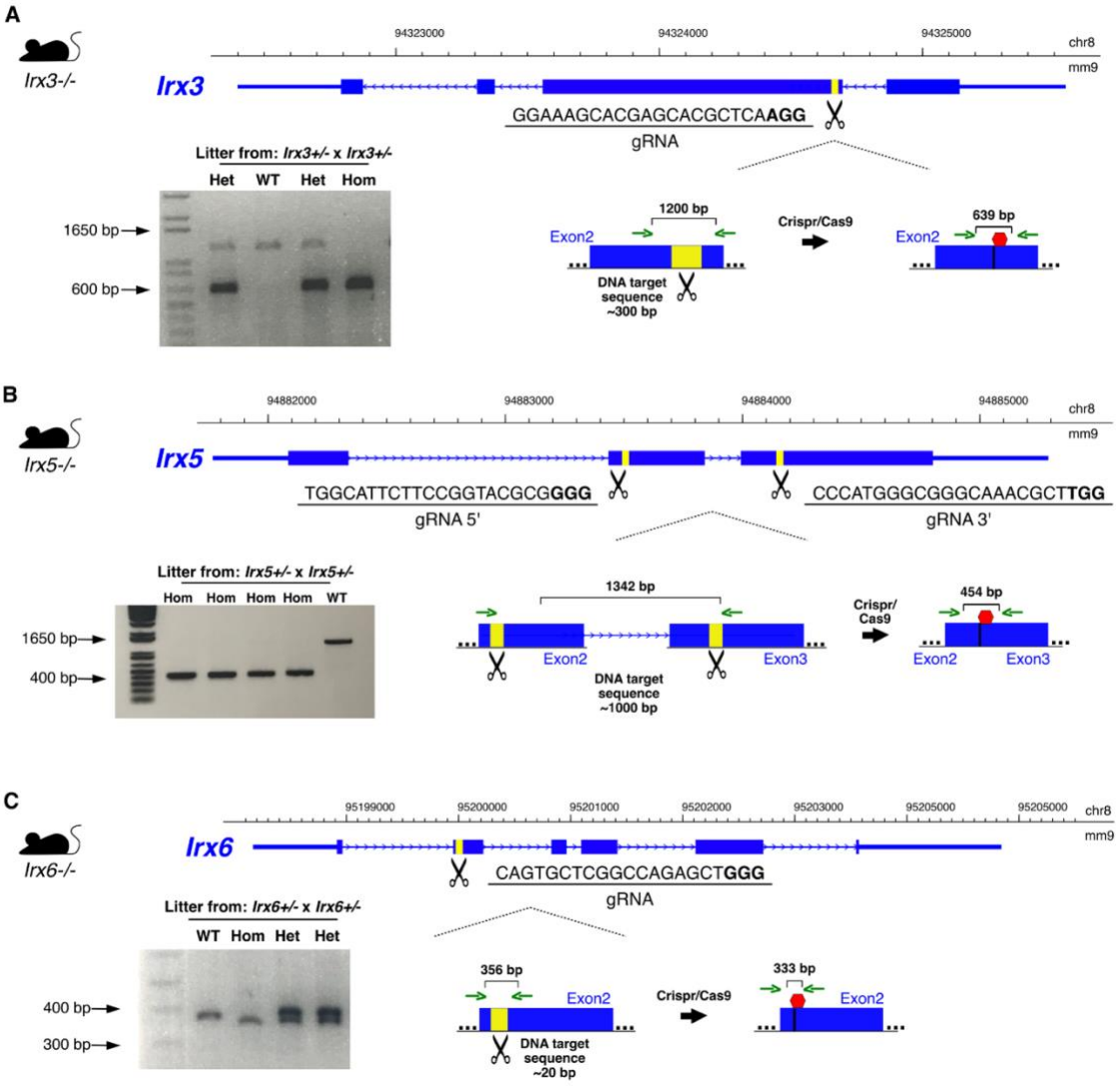

Fig. S3. Generation of *Irx3*, *Irx5*, and *Irx6* mutant mice using CRISPR/Cas9 editing. (**A**) *Irx3*. (**B**) *Irx5*. (**C**) *Irx6*. The gRNA-targeting sequence (gRNA) is underlined, and the PAM sequences are indicated in bold. Exons are represented as thick blue boxes, introns are indicated as blue lines with arrows, and the yellow boxes indicate the DNA-targeting region. Red hexagons indicate a stop codon generating a truncated protein. Agarose gel showing the PCR products generated from DNA containing successfully targeted *Irx3*, *Irx5*, or *Irx6* from F0 mouse tail genomic DNA. The DNA sequence from WT animals was uncleaved and, therefore, larger than the product generated from the mutant.

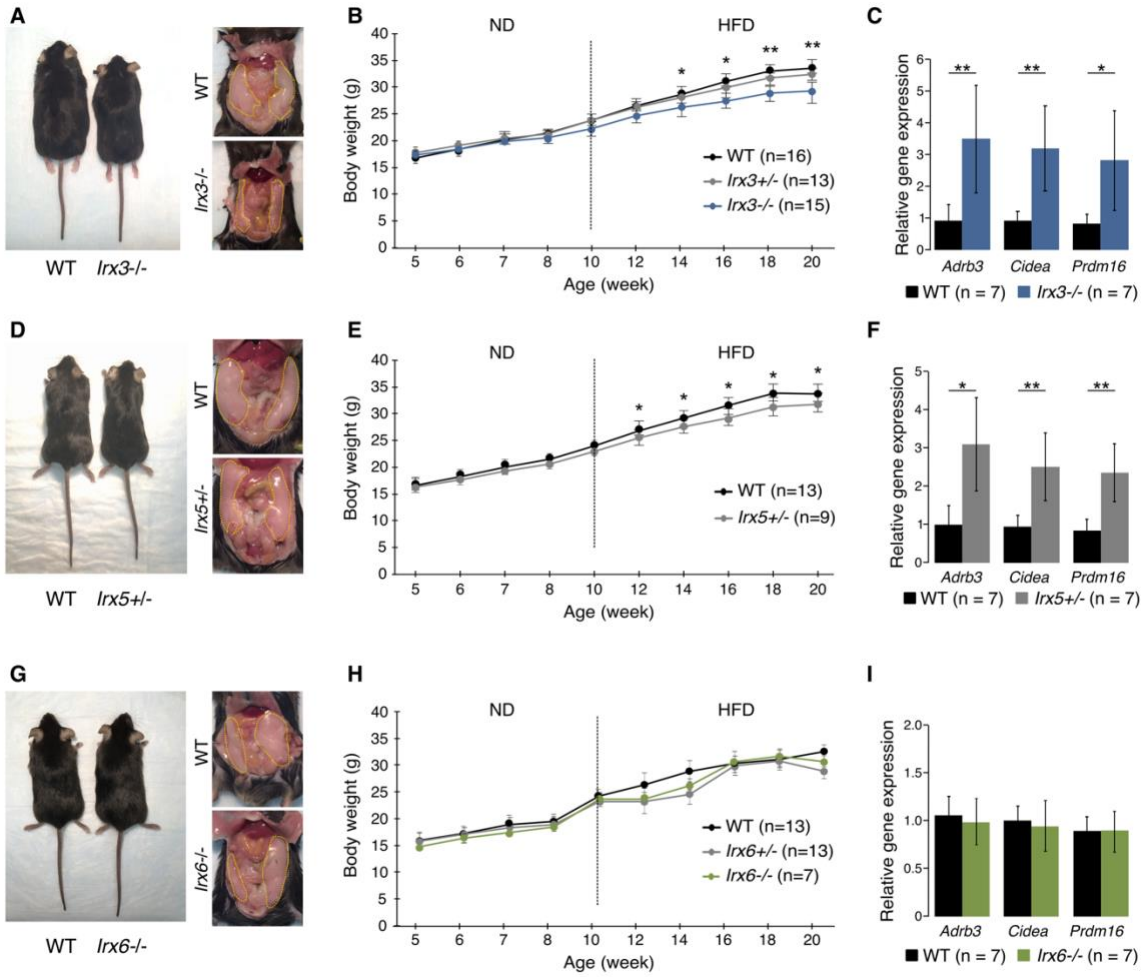

Fig. S4. *Irx3*-/- and *Irx5*+/- knockout mice are lean and protected against diet-induced obesity. (**A, D, G**) Representative photograph (left panel) and anatomic view (right panel) of wild-type (WT) and *Irx3*-/-, *Irx5*+/-, and *Irx6*-/- mice after 10 weeks of being fed a high-fat diet (HFD). Yellow dotted lines show perigonadal white adipose tissue (PWAT). (**B, E, H**) Body weight gain in mice fed a normal diet (ND) and after HFD in *Irx3*-/-, *Irx5*+/-, and *Irx6*-/-, respectively. (**C, F, I**) Relative gene expression in PWAT after HFD comparing WT animals and *Irx3*-/-, *Irx5*+/-, and *Irx6*-/-. Statistical significance was determined by Student t test. Data are expressed as mean ± SEM. **P* < 0.05 and ***P* < 0.01 compared to WT group.

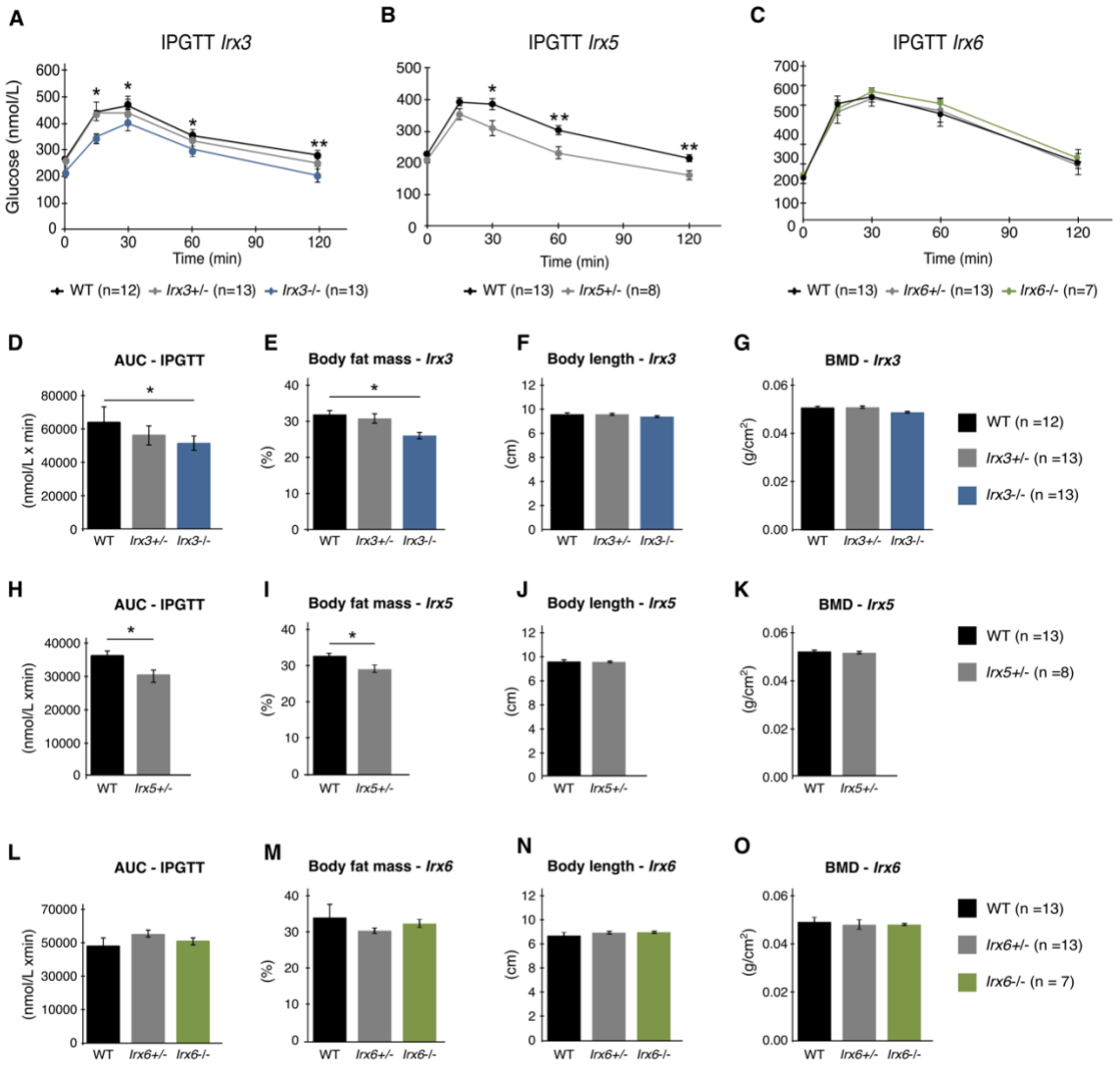

Fig. S5. Effects of a high-fat diet on glucose tolerance, body length, fat mass and bone density. (**A, B, C**) Intraperitoneal glucose tolerance test (IPGTT) in WT and *Irx3*-/-, *Irx5*+/-, and *Irx6*-/- mice, respectively. (**D, H, L**) Graphs show the area under the curve (AUC) of the blood glucose concentration levels measured during IPGTT of WT and *Irx3*-/-, *Irx5*+/-, and *Irx6*-/-, respectively. (**E, I, M)** Body fat mass ratio as a percentage of body weight after HFD for 10 weeks, *Irx3*-/-, *Irx5*+/-, and *Irx6*-/-, respectively. (**F, J, N**) Body length after HFD for 10 weeks, *Irx3*-/-, *Irx5*+/-, and *Irx6*-/-, respectively. (G, K, O) Bone mineral density (BMD) after HFD for 10 weeks, *Irx3*-/-, *Irx5*+/-, and *Irx6*-/-, respectively. Statistical significance was determined by Student *t* test. Data are expressed as mean ± SEM. **P* < 0.05, and ***P* < 0.01 compared to WT group.

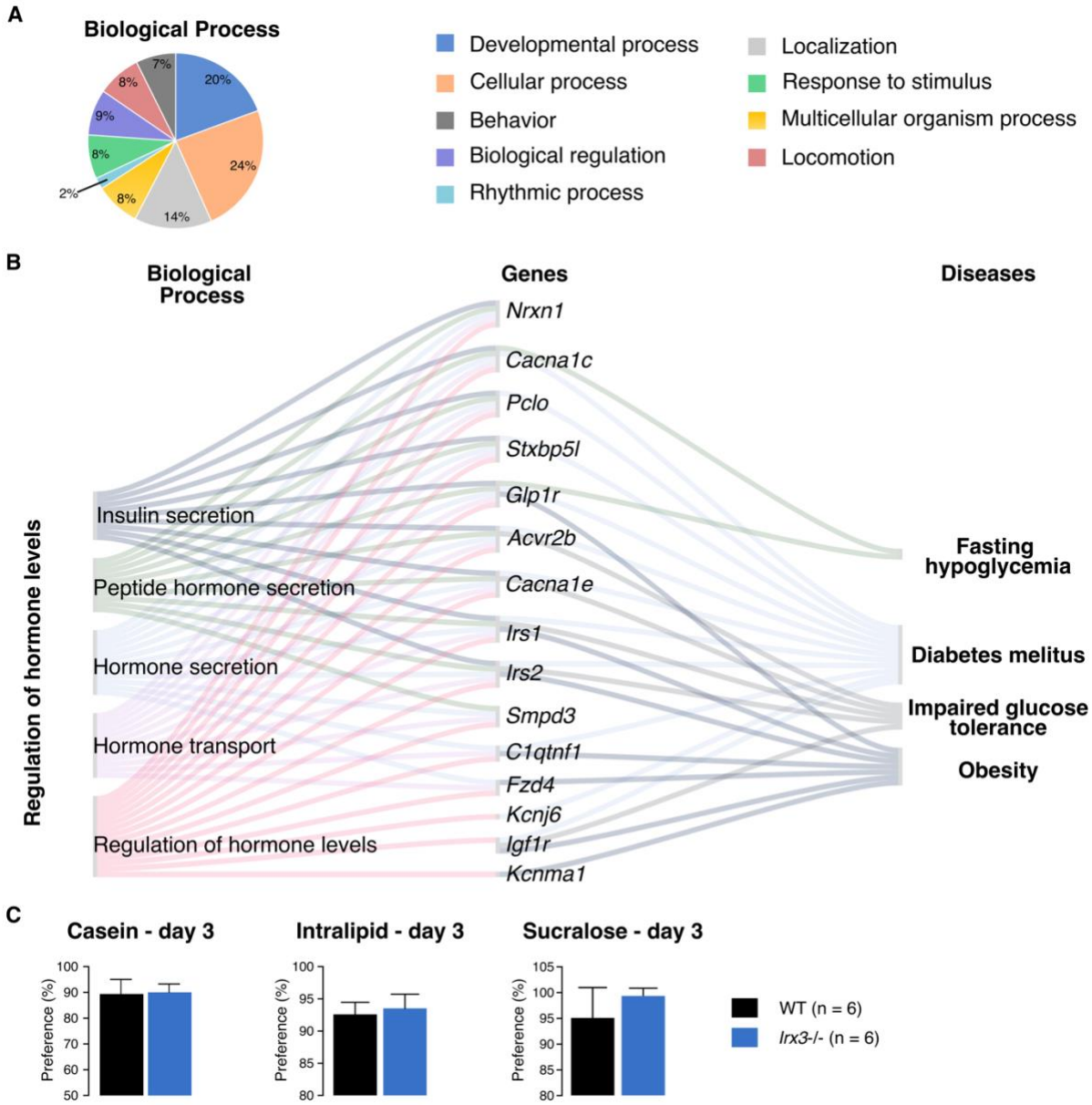

Fig. S6. Expression analysis of *Irx3*-/- hypothalamus and changes in macronutrient selection. (**A**) RNA-seq transcriptional profiling of differential gene expression (DE) between *Irx3*-/- and WT hypothalamus using Gene Ontology (GO). Pie chart shows the distribution of all upregulated biological process in *Irx3*-/- hypothalamus. Numbers indicate the percentages of genes in each GO category. (**B**) Sankey flow diagram showing all genes upregulated in the hypothalami from *Irx3*-/- animals with high enrichment for regulation of hormone levels and the top ranked diseases related to them. Gene symbols are shown. (**C**) Intake of Casein, Intralipid, or Sucralose in WT and *Irx3*-/- mice. Data are expressed as mean ± SEM. **P* < 0.05 compared to WT group.

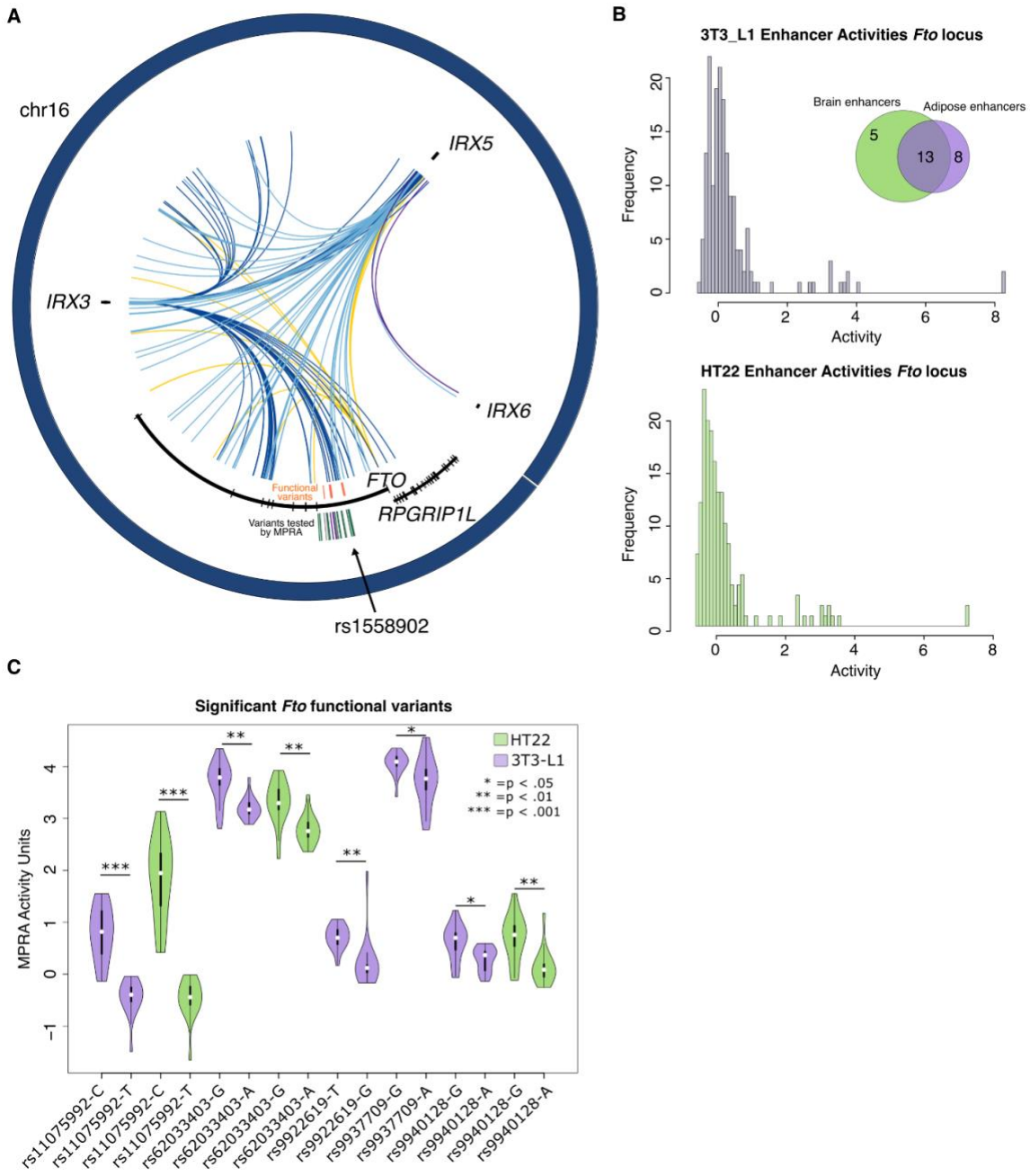

Fig. S7. Several variants within the first intron of *FTO* exhibit allele-specific enhancer activities. (**A**) The genomic landscape of the *FTO* GWAS association locus. Promoter capture Hi-C interactions from SGBS preadipocytes are shown as arcs that originate from the promoter of *IRX3* (dark blue), *IRX5* (light blue), *IRX6* (purple), and *FTO/RPGRIP1L* (yellow). The location of the GWAS lead variant (rs1558902) in the first intron of *FTO* is indicated with an arrow. Variants tested with MPRA mapping within enhancers are shown directly above the arrow as purple, gray, or green colored lines (variants tested by MPRA). Gray variants were significant in both cell types, green were significant in HT22, and purple were significant in 3T3-L1. Variants with allele-specific enhancer properties are colored in orange (functional variants). (**B**) Bar chart of

activity levels for all constructs tested with MPRA, along with a Venn diagram showing the number of enhancers found in each cell type. (**C**) MPRA activity measures for each *FTO* functional variant barcode with significant allelic effects separated by cell type. *q<0.05 two-sided Mann Whitney U test. Haplotype and allele frequencies in the CEU population for variants with allelic effects on enhancer activity (LDHap tool: https://ldlink.nci.nih.gov).

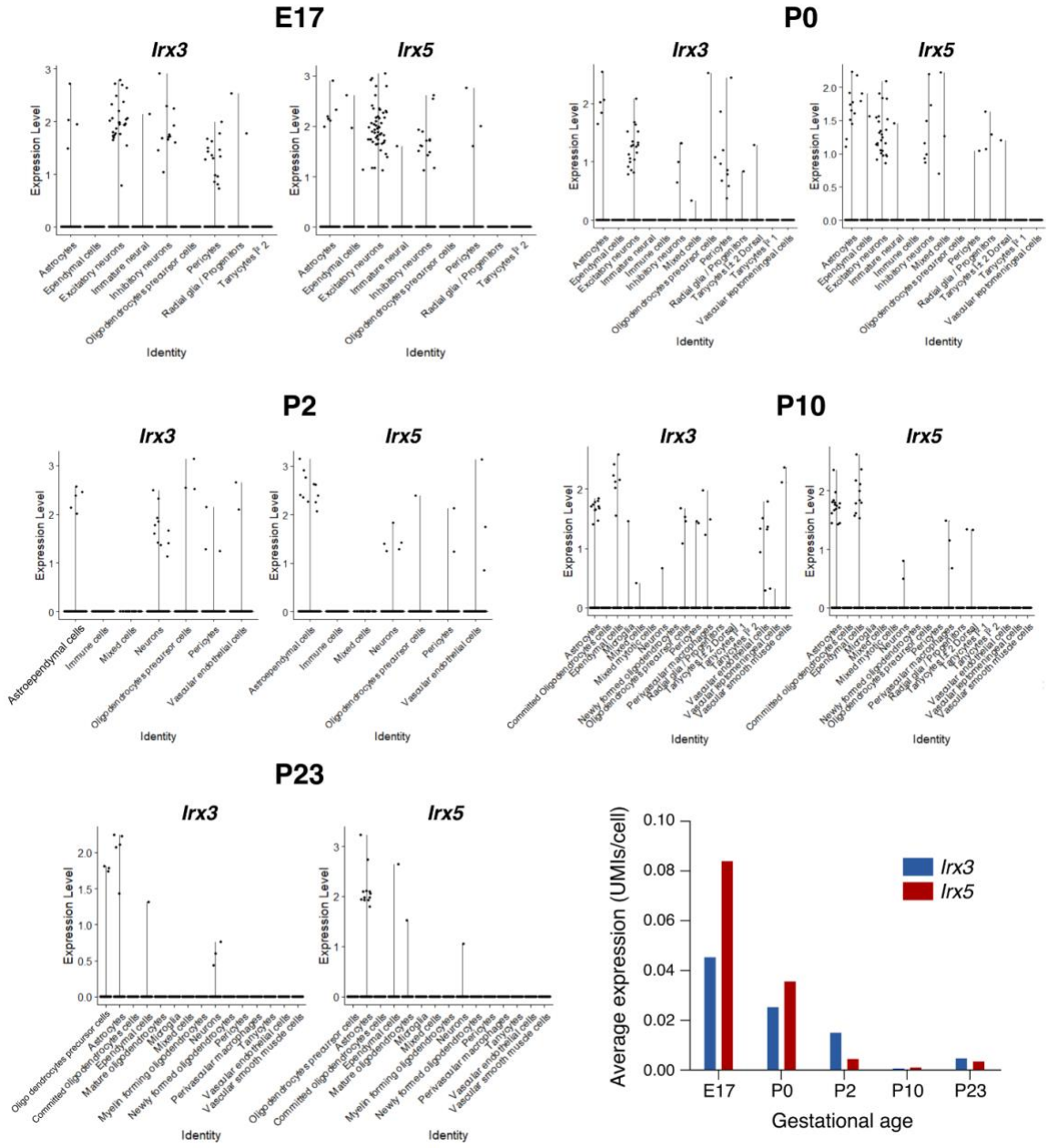

Fig. S8. Expression of *Irx3* and *Irx5* during mouse hypothalamus development. Expression of *Irx3* and *Irx5* in different cell types during hypothalamus development at different stages: E17 (**A**), P0 (**B**), P2 (**C**), P10 (**D**), and P23 (**E**) (*24*). (**F**) Summary of *Irx3* and *Irx5* expression over time during mouse hypothalamus development. Unique molecular identifiers (UMIs).

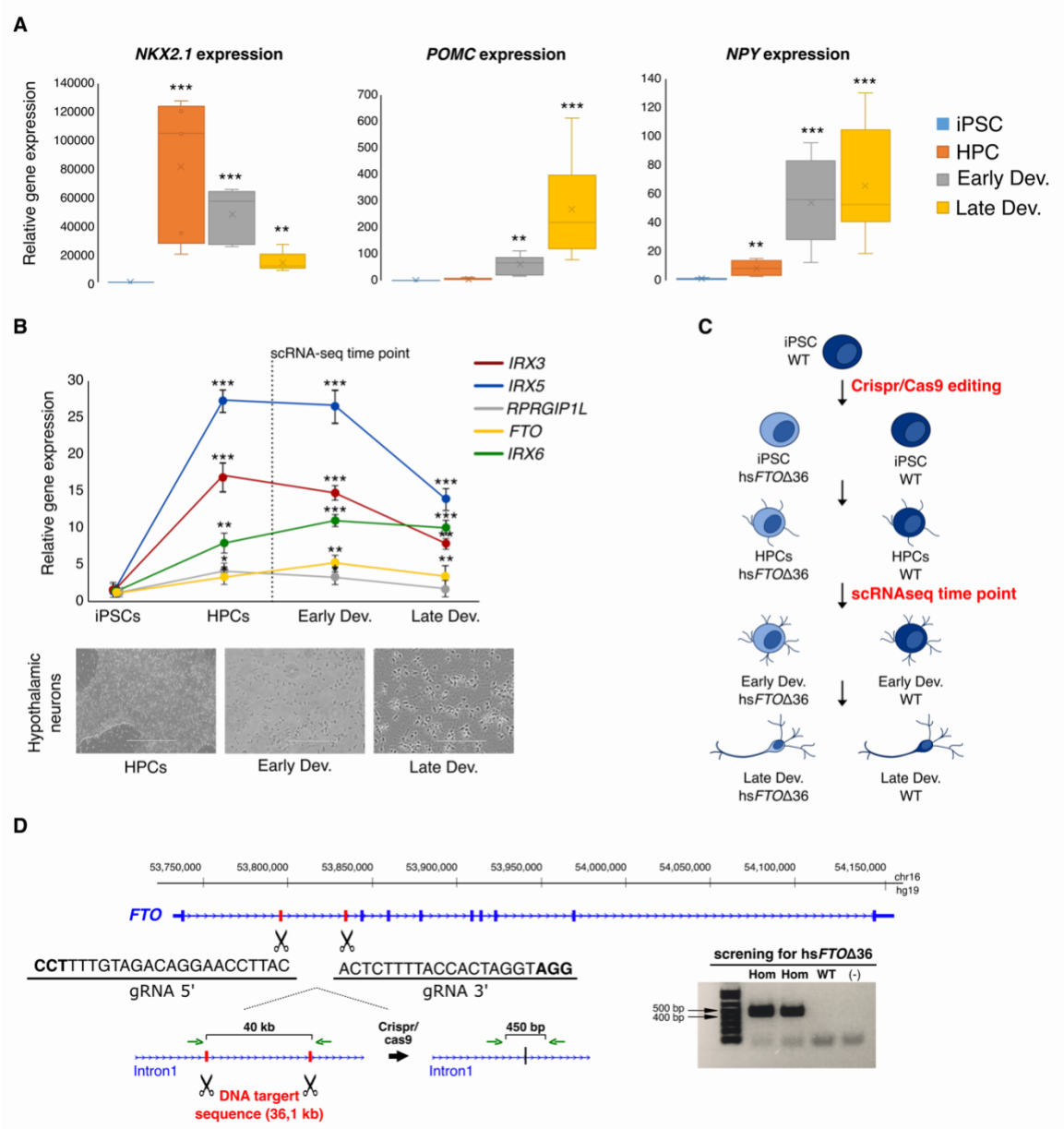

Fig. S9. Generation of human hypothalamic arcuate-like neurons from iPSCs. (**A**) Bulk qPCR analysis of hypothalamic markers at four different time points during the differentiation of hypothalamic neurons. (**B**) qPCR analysis of *IRX3*, *IRX5*, *IRX6*, *FTO*, and *RPGRIP1L* expression at four different time points during the differentiation of hypothalamic neurons. The dashed line indicates the time point when the scRNA-seq assay was performed. Representative microscopy image of each differentiation stage is shown. Bars, 200 µm (**C**) Schematic diagram illustrating the different stages of hypothalamic ARC-like neuron generation from iPSCs. (**D**) Overview of the CRISPR/Cas9 strategy to delete 36,100 bp encompassing the *FTO* obesity-associated locus in human iPSC. The gRNA-targeting sequence (gRNA) is underlined, and the PAM sequences are indicated in bold. Exons are represented as thick blue boxes, introns are indicated as blue lines with arrows, and the red boxes indicates the DNA-targeting

region. Agarose gel shows the PCR products generated from DNA containing the successfully targeted *FTO* obesity-associated locus from iPSC genomic DNA. The abbreviations are: iPSCs (induced pluripotent stem cells); HPCs (hypothalamic progenitor cells); Early Dev (hypothalamic neurons at early development time point); Late Dev (hypothalamic neurons at late development time point). For qPCR analysis, the charts show normalized expression relative to iPSCs. Error bars represent the ±SEM from four biological replicates. $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$.

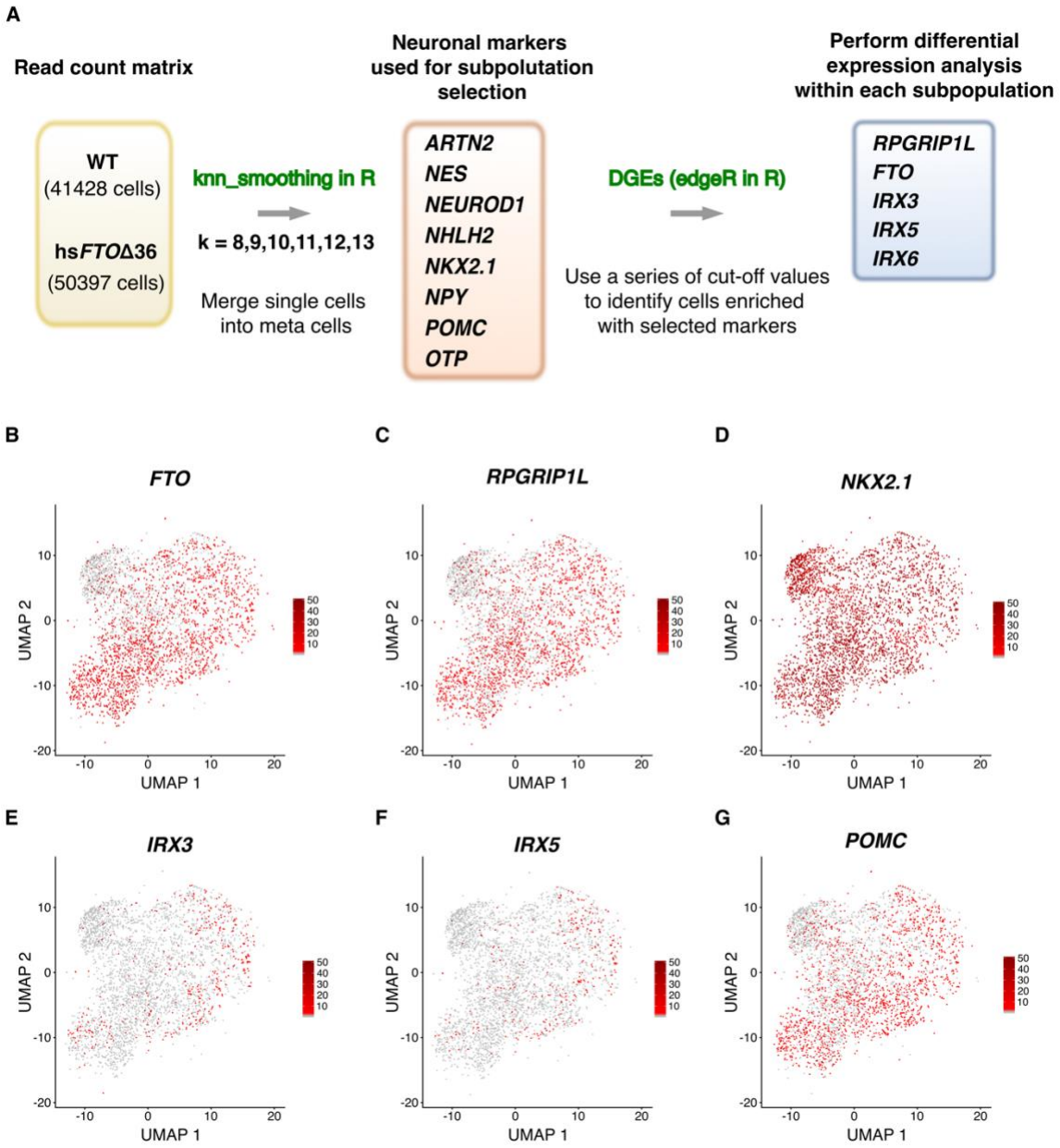

Fig. S10. Flowchart of scRNA-seq data analyses and UMAP embedding of merged scRNA-seq profiles from WT and hs*FTO*Δ36 human hypothalamic arcuate-like neurons. (**A**) Flowchart displaying the analytical pipeline used to evaluate the impact of deleting the obesity-associated interval on the expression of genes in developing hypothalamic neurons. (**B-G**) The UMAP plots show the expression of indicated genes: *FTO*, *RPGRIP1L*, *NKX2.1*, *IRX3*, *IRX5*, and *POMC*, respectively, along UMAP_1 and UMAP_2. Random subsampling of 3500 of total 91825 cells from the KNN-aggregated dataset of single cells with K = 10.

Table S5. Functional variants with significant differences in enhancer activity between alleles identified by MPRA and PMCA

Lead SNP	Candidate variant (dbSNP)	Genomic Coordinate
rs1558902	rs11075992 ^ *	chr16:53,819,978-53,820,153
rs1558902	rs62033403 ^	chr16:53,822,149-53,822,324
rs1558902	rs9922619 ^ *	chr16:53,831,683-53,831,858
rs1558902	rs9937709 ^	chr16:53,820,725-53,820,900
rs1558902	rs9940128 ^	chr16:53,800,666-53,800,841
rs1558902	rs11642015 ♀ *	chr16:53,800,870-53,801,045
rs1558902	rs14211085 ♀ *	chr16:53,802,386-53,802,561

^ indicates variant identified by MPRA

♀ indicates variant identified by PMCA

* indicates variants validated with luciferase assay

Table S9. Marker genes defining cell clusters.

Cluster name	**Genes**
Pluripotency	*SOX2, NANOG and OCT4*
Hypothalamic progenitor cells (HPC)	*HSP5, TAF7* and *SRSF2*
Hypothalamic neurons at an early developmental time point (Early Dev)	*NES, SOX3, NEUROG3, ASCL1, HES1, CCND1, CDK4, DLL1, DLL3*
Hypothalamic neurons at a late developmental time point (Late Dev)	*DLX1, DLX2, DLX5, DLX6, SOX11, SOX12 OTP, NEUROD1, STMN1/2, Nkx2-1, NPY, NHLH2, ARNT2, ISL1*
Radial Glia	*VIM*
Astrocytes	*GLAST, BLBP, APOE, HSPA5*

Table S10. Oligonucleotides and gRNA target sites used in this study.

gRNA target sites used to generate transgenic mice		
Target region	sgRNA (1)	sgRNA (2)
Irx3-/-	GGAAAGCACGAGCACGCTCAAGG	N/A
Irx5-/-	TGGCATTCTTCCGGTACGCGGGG	CCCATGGGCGGGCAAACGCTTGG
Irx6-/-	CAGTGCTCGGCCAGAGCTGGG	N/A
mm*Fto*Δ20	GTTAAAATACCCTCTCATCTGGG	GGTCCTTGGGATCTCCGCTTTGG
crRNA target sites used to generate hsFTOΔ36 cell line		
Region	crRNA (1)	crRNA (2)
hs*FTO*Δ36	CCTTTTGTAGACAGGAACCTTAC	ACTCTTTTACCACTAGGTAGG
Product length and primers sequences used to genotype transgenic mice and iPSC		
Gene	Primer 5 -3 (sense)	Primer 5 -3 (antisense)
Irx3	TACCAGTACATCCGCCCTCT	TGACCAGATCTTGGGCTTCT
Irx5	AAGGACAGAGGGAAAGTGGG	TCTTTAGCCAAAACGTCCGC
Irx6	CCTTGCATAATCTCCTCTGG	GCCCCATTCCATTACTTACC
mm*Fto*Δ20	CCCAATTACTTTTGGCCTTTT	CTCCCAGGCATAAAATTCA
hs*FTO*Δ36	GGGAGGAGAGTGTGGGTTTT	GTGAAGGCTTTGGATTTGGAG

4C-seq PCRs oligos			
Gene promoter	Reading primer	Non-reading primer	Enzymes
Irx3	CTCCAGTTTTGACCCCCATG	GAATCTCGGCCAATCTGTTC	NlaIII/DpnII
Irx5	CGATGCCTGCTTCTCATG	GGCATCCATCGACTCTCC	NlaIII/DpnII
Irx6	ATTCGGTGGCCAGGGTAC	CGAGTTCTGCGATTGGTAAG	Csp6I/NlaIII
Fto	TGGAAGTAGAGGCAGGGAG ATC	GACTCCAGGCTCAGTTGTGA	DpnII/Csp6I

Table S11. Fosmid clones used as FISH probes.

Region	Whitehead fosmids http://bacpac.chori.org/download/wi1_mm9/chr8.bed	Mm9 coordinates	Probe length (bases)
FTO adjacent LD block	WIBR1-0590M09	chr8:93911541-93950923	39383
Fto/Rpgrip1l promoter	WIBR1-0983O10	chr8:93807527-93849628	42102
Irx3 gene	WIBR1-1206E19	chr8:94307305-94341470	34166
Irx5 gene	WIBR1-1060D04	chr8:94865335-94904882	39548
Irx6 gene	WIBR1-0901O22	chr8:95190971-95221152	30182
FTO LB block del.	WIBR1-0214P05	Chr8:93875571-93913542	37972

Table S12. Twenty-nine genomic loci identified that passed criteria for genome-wide association significance.

Index SNPs for regions with $P < 5x10^{-8}$ are reported. Position – SNP position according to the NCBI Genome Reference Consortium human genome build 37 (GRCh37); Alleles – effect/no effect on genomic reference strand; EAF – effect allele frequency across all study participants; OR – odds ratio for the high-risk allele; CI – 95% confidence interval; Nearest gene – nearest gene to the index SNP

rsID	Chr	Position	Alleles (effect/no effect)	EAF	OR	*P* value	Nearest gene
rs838133	19	49259529	A/G	0.44	1.146 (1.128 - 1.165)	1.98E-49	*FGF21*
rs1421085	16	53800954	C/T	0.41	1.097 (1.079 - 1.115)	3.62E-23	*FTO*
rs6980548	8	64682296	A/G	0.25	1.088 (1.068 - 1.108)	1.78E-15	*YTHDF3*
rs6804842	3	25106437	A/G	0.42	1.075 (1.057 - 1.092)	6.90E-15	*RARB*
rs7924535	11	126754032	G/A	0.49	1.070 (1.053 - 1.088)	1.08E-13	*KIRREL3*
rs35332062	7	73012042	A/G	0.12	1.120 (1.089 - 1.152)	6.85E-13	*MLXIPL*
rs11940694	4	39414993	A/G	0.41	1.080 (1.059 - 1.100)	7.25E-13	*KLB*
rs7067170	X	68382836	G/A	0.75	1.064 (1.047 - 1.080)	1.20E-12	*PJA1*
rs9620391	22	24875123	T/C	0.39	1.068 (1.050 - 1.085)	2.37E-12	*UPB1*
rs4940992	18	58835333	C/T	0.29	1.072 (1.053 - 1.091)	4.55E-12	*CDH20*
rs7849833	9	23198470	G/A	0.37	1.065 (1.047 - 1.083)	2.64E-11	*ELAVL2*
rs1009468	9	15825230	A/G	0.54	1.063 (1.046 - 1.080)	2.96E-11	*CCDC171*

rs4784140	16	61643216	A/G	0.78	1.074 (1.053 - 1.095)	8.59E-11	*CDH8*
rs4079124	5	50884221	G/T	0.68	1.065 (1.047 - 1.084)	9.63E-11	*ISL1*
rs1805415	1	226570840	C/T	0.84	1.084 (1.060 - 1.108)	1.70E-10	*PARP1*
rs6016783	20	35556192	T/G	0.61	1.062 (1.044 - 1.079)	1.79E-10	*SAMHD1*
rs6475657	9	22514381	G/T	0.41	1.061 (1.044 - 1.079)	1.86E-10	*DMRTA1*
rs1442883	2	59970660	C/A	0.77	1.068 (1.048 - 1.089)	1.35E-09	*BCL11A*
rs949824	12	63482910	C/T	0.68	1.060 (1.042 - 1.079)	2.14E-09	*AVPR1A*
rs3809863	17	45385012	T/C	0.47	1.056 (1.039 - 1.073)	2.42E-09	*ITGB3*
rs2977332	8	76754934	C/T	0.68	1.059 (1.041 - 1.078)	4.01E-09	*HNF4G*
rs781736	6	125201001	A/G	0.46	1.054 (1.037 - 1.071)	1.32E-08	*NKAIN2*
rs2582045	5	87596675	T/G	0.52	1.053 (1.036 - 1.070)	1.94E-08	*TMEM161B*
rs2851507	7	69115202	A/G	0.35	1.055 (1.038 - 1.073)	1.96E-08	*AUTS2*
rs780094	2	27741237	C/T	0.59	1.054 (1.036 - 1.071)	2.30E-08	*GCKR*
rs13198474	6	25874423	A/G	0.06	1.109 (1.073 - 1.146)	3.10E-08	*SLC17A3*
rs8095324	18	24131659	A/G	0.6	1.053 (1.036 - 1.071)	3.15E-08	*KCTD1*
rs472418	18	58187465	A/C	0.33	1.056 (1.037 - 1.074)	3.32E-08	*MC4R*

rs4475363	6	8905429	T/C	0.91	1.092 (1.061 - 1.123)	3.45E-08	*SLC35B3*

Additional Data (available as an excel (.xlsx) file online)

Table S1. GO/ToppGene analysis.

This Excel file contains a list of all up-regulated genes within cellular metabolic processes and regulation of hormone levels GO categories. Also, a list of the top 50 ranked diseases associated with these genes.

Table S2. Sweet vs. salty preference GWAS summary statistics.

This excel file contains the summary statistics for a GWAS performed on sweet vs. salty preference data collected by 23andMe, including SNP information, unadjusted P values, genomic control adjusted P values, and genotype counts

Table S3. Variants tested in the Massively Parallel Reporter Assay.

This Excel file contains all 87 SNPs tested in the MPRA assay and the functional variants with enhancer activity in at least three replicates tested in each cell line (3T3-L1 and HT22).

Table S4. Significant FTO locus functional variants identified using MPRAs.

This file contains SNPs that were identified as statistically significant enhancer modulating variants using a massively parallel reporter assay performed in HT22 and/or 3T3-L1 cell lines

Table S6. PMCA accessible variants.

This Excel file includes the variant by PMCA analysis with high score located in functionally conserved regions. Columns define the metrics which contribute to the PMCA overall score as defined in (*21*), with Sall in the restricted set of binding sites representing the log10 strength of conservation. W(TFBS), number of transcription factor (TF) binding sites; p-est (TFBS), estimated probability of the number of TF binding sites in the shuffled control; W(modules), number of TF binding modules; p-est(modules), estimated probability of the number of TF binding modules in the shuffled control; W (TFBS_in_modules), number of TF binding sites contained in modules; p-est (TFBS_in_modules), estimated probability of the number of TF binding sites in the shuffled control. These metrics are repeated both across all species and only in the restricted set (present across at least half the species).

Table S7. DeepSEA predictions.

This file contains all SNPs predicted by DeepSEA analysis with score < 0.01. Functional significance score, the -log10 p-value of a given SNP being functional, as defined in the DeepSEA manuscript (*22*).

Table S8. Cell type specific Basset model predictions.

This file contains all SNPs tested in the Basset model comparing the allelic activity across 170 different cell types and tissues. We focus on the two cell types used in this study, preadipocytes (adipocyte_d0, ATAC-seq from immortalized adipose derived mesenchymal stem cells at day 0 of differentiation toward adipocytes) and hypothalamic neurons (Hypothalamic neurons, ATAC-seq from iPSC-derived hypothalamic arcuate-like neurons). Hypothalamic_Neurons_Strong represents the hypothalamic neuron data subset to the strongest peaks (those with a -log10 q-value > 6).

Table S13. Number of cells and genes retained for the sc-RNA-seq analysis.

This Excel file contains the number of retained cells and genes after the initial single cell RNA-seq pre-processing, and the mitochondrial gene content for each replicate of the single cell RNA-seq analysis. Threshold of 20% for mitochondrial gene content was applied.

Table S14. Summary statistics for *POMC*-enriched cell datasets.

This Excel file contains the number of cells across the WT and hs*FTO*Δ36 conditions and the mitochondrial content for each *POMC* cutoff and KNN K value.

Table S15. Result of DE analysis for *POMC*-enriched cells.

This Excel file contains the list of the logFC and adjusted *P* values corresponding to different cutoffs and different Ks for cells expressing *POMC*.

Table S16. Summary statistics for cells expressing seven different neuronal marker genes.

This Excel file contains the number of cells across the WT and hs*FTO*Δ36 conditions and the mitochondrial content for each *POMC* cutoff and KNN K value. Neuronal markers: *ARNT2, NES, NEUROD1, NHLH2, NKX2.1, NPY,* and *OTP*.

Table S17. Result of DE analysis for cells enriched for other marker genes.

This Excel file contains the list of the logFC and adjusted *P* values corresponding to different cutoffs and different Ks for cells expressing different neuronal markers: *ARNT2, NES, NEUROD1, NHLH2, NKX2.1, NPY,* and *OTP*.