

Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*

Scott Smemo^{1*}, Juan J. Tena^{2*}, Kyoung-Han Kim^{3*}, Eric R. Gamazon⁴, Noboru J. Sakabe¹, Carlos Gómez-Marín², Ivy Aneas¹, Flavia L. Credidio¹, Débora R. Sobreira¹, Nora F. Wasserman¹, Ju Hee Lee³, Vijitha Puviindran³, Davis Tam³, Michael Shen¹, Joe Eun Son⁵, Niki Alizadeh Vakili³, Hoon-Ki Sung⁵, Silvia Naranjo², Rafael D. Acemel², Miguel Manzanares⁶, Andras Nagy⁵, Nancy J. Cox^{1,4}, Chi-Chung Hui³, Jose Luis Gomez-Skarmeta² & Marcelo A. Nóbrega¹

Genome-wide association studies (GWAS) have reproducibly associated variants within introns of *FTO* with increased risk for obesity and type 2 diabetes (T2D)^{1–3}. Although the molecular mechanisms linking these noncoding variants with obesity are not immediately obvious, subsequent studies in mice demonstrated that *FTO* expression levels influence body mass and composition phenotypes^{4–6}. However, no direct connection between the obesity-associated variants and *FTO* expression or function has been made^{7–9}. Here we show that the obesity-associated noncoding sequences within *FTO* are functionally connected, at megabase distances, with the homeobox gene *IRX3*. The obesity-associated *FTO* region directly interacts with the promoters of *IRX3* as well as *FTO* in the human, mouse and zebrafish genomes. Furthermore, long-range enhancers within this region recapitulate aspects of *IRX3* expression, suggesting that the obesity-associated interval belongs to the regulatory landscape of *IRX3*. Consistent with this, obesity-associated single nucleotide polymorphisms are associated with expression of *IRX3*, but not *FTO*, in human brains. A direct link between *IRX3* expression and regulation of body mass and composition is demonstrated by a reduction in body weight of 25 to 30% in *Irx3*-deficient mice, primarily through the loss of fat mass and increase in basal metabolic rate with browning of white adipose tissue. Finally, hypothalamic expression of a dominant-negative form of *Irx3* reproduces the metabolic phenotypes of *Irx3*-deficient mice. Our data suggest that *IRX3* is a functional long-range target of obesity-associated variants within *FTO* and represents a novel determinant of body mass and composition.

Noncoding variation in single nucleotide polymorphisms (SNPs) within a 47-kilobase (kb) region of high linkage disequilibrium in introns 1 and 2 of *FTO* remains the strongest genetic association with risk to polygenic obesity in humans^{1–4}. Individuals homozygous for risk alleles of the associated SNPs weigh approximately 3 kg more than individuals homozygous for non-risk alleles, underscoring the significant phenotypic impact of these common variants². Multiple follow-up studies directly implicated *FTO* as a gene controlling body mass and composition. Null *Fto* alleles in mice result in lean phenotypes, with reciprocal models overexpressing *FTO* displaying increased body weight^{4–6}. However, there is no direct evidence that enhancers within the obesity-associated *FTO* introns are connected with regulation of *FTO* expression. Expression quantitative trait locus (eQTL) analyses have systematically failed to show association between the obesity-associated SNPs and *FTO* expression in human tissues^{7–9}.

To chart directly the *cis*-regulatory circuitry within the *FTO* locus, we used circular chromosome conformation capture followed by high-throughput sequencing (4C-seq). We carried out 4C-seq in whole mouse embryos (embryonic day 9.5; E9.5) and in adult (8 weeks) mouse brains, as previous work suggests that brain *FTO* expression modulates

metabolic parameters^{6,10}. We profiled the genomic interactions with promoters of genes located within a 1-megabase (Mb) window around the obesity-associated SNPs (mouse genome build mm9, chromosome 8: 93,725,000–94,725,000), including *Fto* and *Rpgrip1l*, and *Irx3*, a half-megabase downstream. There was a clear difference in interaction patterns among genes (Fig. 1a). The *Fto* promoter chiefly participates in genomic interactions proximal to the gene promoter. Although in mouse embryos these interactions include the obesity-associated intronic region (Fig. 1b), no such interactions are detected in adult mouse brains (Extended Data Fig. 1 and Supplementary Table 1). In marked contrast, the promoter of *Irx3* participates in numerous long-range interactions across a broad genomic region encompassing nearly 2 Mb, including robust interactions with the obesity-associated interval within *FTO* (Fig. 1a). We confirmed these interactions between *Irx3* and the *Fto* obesity-associated region using chromatin conformation capture (3C) in adult mouse brains (Extended Data Fig. 2c). These data suggest that the obesity-associated interval is likely to be part of the regulatory landscape of *Irx3*. We next inferred that the long-range interactions between the obesity-associated *FTO* intron and *IRX3* represent a conserved feature in vertebrate genomes. Inspecting the human ENCODE data set, we observe that human breast cancer MCF-7 cells display the same pattern of long-range looping between the obesity-associated interval and *IRX3*, but not *FTO*, assayed by chromatin interaction analysis with paired-end tag sequencing (ChIA-PET) (Extended Data Fig. 2a). This is corroborated further by recent Hi-C (high-throughput 3C) data in human fibroblasts¹¹ (Extended Data Fig. 2b). We also observed a similar pattern of chromatin interactions in zebrafish embryos, assayed by 4C-seq (Extended Data Fig. 1c). Together, these data suggest that the obesity-associated *FTO* intron is mediating functional interactions with *IRX3* in the human, mouse and zebrafish genomes.

After revealing evidence of direct looping between *FTO* intronic regions and *IRX3*, we next showed that enhancers within the obesity-associated interval possess functional characteristics compatible with *IRX3* expression. We observed that the degree of evolutionary conservation in noncoding sequences of the *FTO*-*IRX* locus (chromosome 16: 53,731,249–54,975,288) places it on the top 2% of similarly sized genomic intervals, suggesting the presence of many functional noncoding elements spread over the region (see Methods). More specifically, ENCODE data suggest that the 47-kb obesity-associated interval is fraught with *cis*-regulatory elements, evidenced by an abundance of enhancer-associated chromatin marks, DNase hypersensitive sites, and transcription-factor binding events across this genomic region (Fig. 1b). We tested three human DNA fragments from the 47-kb obesity-associated region for their putative enhancer properties using an *in vivo* mouse reporter assay. These fragments overlap enhancer-like chromatin marks in multiple cell lines. All three fragments displayed enhancer activity in neonatal (postnatal day

¹Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. ²Centro Andaluz de Biología del Desarrollo (CABD), Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, Carretera de Utrera Km1, Sevilla 41013, Spain. ³Program in Developmental & Stem Cell Biology, The Hospital for Sick Children, and Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada. ⁴Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA. ⁵Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5T 3H7, Canada. ⁶Cardiovascular Development and Repair Department, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Madrid 28029, Spain.

*These authors contributed equally to this work.

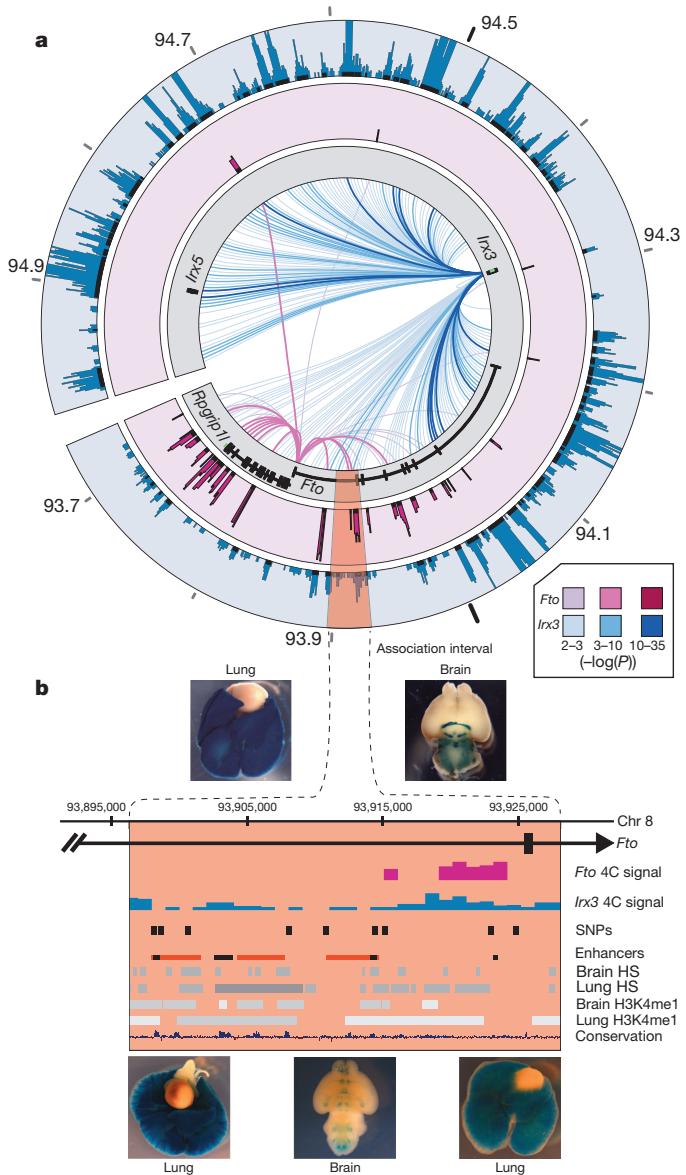


Figure 1 | Long-range interactions in the *IRX3*-*FTO* locus. **a**, Mouse embryo 4C-seq interactions emanating from each promoter are displayed as links across the circle (darker link implies greater significance). Outer plots show significance of interactions above background ($-\log(P)$ value). The obesity-associated interval is highlighted red. **b**, Magnified view of the association interval. Contained are the orthologous locations of obesity-associated SNPs (black pips), and epigenetic marks associated with regulatory elements. Endogenous *Irx3* expression is shown in lung and brain in *Irx3*^{lacZ} knock-in mouse (top two images). Three enhancers drive reporter expression in lungs and brain (rectangles in the Enhancers row, corresponding to bottom three images). Other enhancers¹⁷ are shown in black. HS, DNase hypersensitive sites.

0 (P0)-P1) mice. Interestingly, two of the fragments drove strong reporter gene expression in lungs (Fig. 1b). *Irx3* is highly expressed during lung development (Fig. 1b)^{12,13}, whereas *Fto* expression in lungs is limited (Extended Data Fig. 3a and refs 2, 14–16). Consistent with our data, a previous report has shown that the *FTO* obesity-associated region contains enhancers that drive *Irx3*-like expression in zebrafish¹⁷.

Our data suggest that the regulatory landscape of *IRX3* spreads over megabase distances, whereas *FTO* expression is primarily regulated by regions proximal to its promoter. To test this, we engineered a human bacterial artificial chromosome (BAC) spanning 162 kb of the *FTO* locus, including its promoter and the 47-kb obesity-associated region (Extended Data Fig. 3). We recombined a reporter cassette at the

FTO translation start, and generated transgenic mice harbouring the engineered BAC. Transgenic mice expressed the reporter gene in multiple tissues, recapitulating the endogenous expression pattern of *FTO*¹⁴. We next determined that a 1.2-kb region corresponding to the *FTO* promoter is sufficient to recapitulate most of the *FTO* expression pattern (Extended Data Fig. 3). In contrast, a 2.8-kb region corresponding to the *IRX3* promoter does not recapitulate any of the endogenous expression patterns of *IRX3*, suggesting that *IRX3* expression relies on long-range *cis*-regulatory elements. A total of 15 human sequences in the broader *FTO*-*IRX3* locus have been characterized as *in vivo* enhancers in mouse^{17,18}, driving reporter expression in *Irx3*-expressing tissues, including eye, limb, brain, neural tube and branchial arches^{12,19} (Supplementary Table 2). Although these enhancers lie outside the obesity association interval, they are consistent with our chromatin looping data pointing to a broad *Irx3* regulatory landscape, extending into *Fto*. Further supporting these findings, the genomic interval spanning *FTO* and *IRX3* has been proposed to be part of a single genomic regulatory block, based on the presence of extended synteny across deep phylogenies and a high density of highly conserved noncoding elements¹⁷, as well as patterns of CTCF binding and chromatin interactions²⁰. Under these models, our data support the idea that the broad expression patterns of *FTO* are primarily regulated by elements proximal to the promoter, and that *IRX3* is endowed with an ancient, extensive *cis*-regulatory circuitry extending into *FTO*.

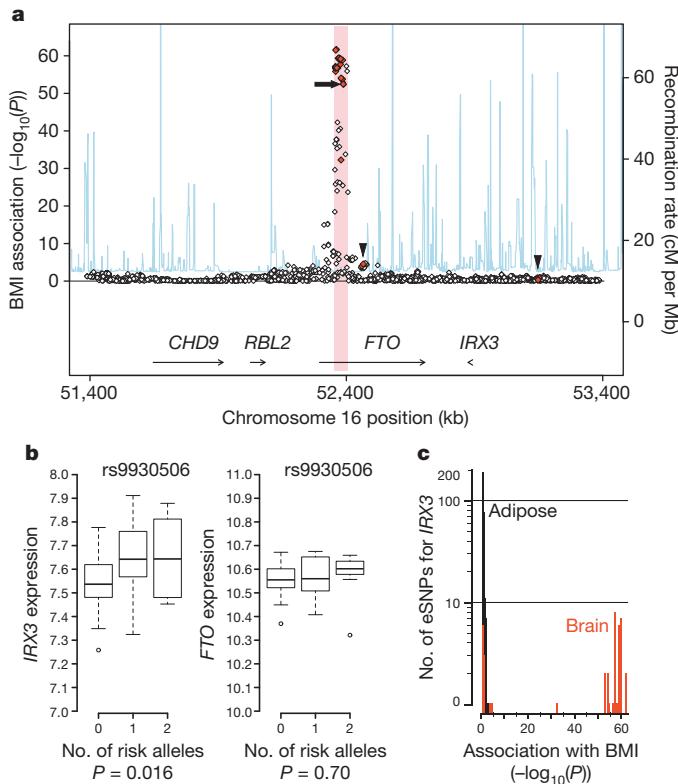


Figure 2 | BMI-associated SNPs are associated with expression of *IRX3*, but not *FTO*, in human brain. **a**, SNPs associated with BMI in the *FTO* locus. Forty-three SNPs associated with *IRX3* expression (eSNPs) are shown in red, including 11 also associated with BMI. No SNPs are associated with *FTO* expression ($P > 0.05$). Thick arrowheads, *IRX3* eSNPs outside the obesity-associated region (pink); thick arrow, rs9930506. **b**, In cerebellum, the allele of rs9930506 associated with increased BMI (risk allele) is correlated with increased *IRX3* expression and not with *FTO* expression. Vertical middle bars represent the first and third quartiles, horizontal lines represent the median, and open circles denote outliers beyond 1.5 times the interquartile range. **c**, Histogram plotting the number of SNPs associated with *IRX3* expression in adipose (black) and brain (red) (y axis) by the significance of their association with BMI (x axis). Full details of statistics are provided in the Methods.

We next determined that gene expression studies in human brains corroborate our chromatin looping data, showing that the obesity-associated SNPs are associated with expression levels of *IRX3*, but not *FTO*. Our data and that of others¹⁷ (Fig. 1b) demonstrate that the obesity-associated SNPs overlap several *in vivo* enhancers, raising the possibility that allelic variants may disrupt enhancer activity and alter expression of their target gene(s). To test this, we carried out eQTL mapping in human brain samples. *FTO* and *IRX3* are highly expressed in multiple regions of the brain including cerebellum and hypothalamus (Extended Data Fig. 4). Using a data set of 153 brain samples from individuals of European ancestry²¹, represented by cerebellum, we found significant association between 11 SNPs previously associated with increased body mass index (BMI), and expression of *IRX3*, but not *FTO* (Fig. 2a and Supplementary Table 3). These common SNPs have minor allele frequencies ranging from 42.5 to 47.5% and are in strong linkage disequilibrium with each other (Extended Data Fig. 5). In all cases, such as for SNP rs9930506 (Fig. 2b), the allele associated with increased BMI is also associated with increased *IRX3* expression. Further analysis, using the GIANT Consortium data set of SNPs associated with BMI in 249,796 individuals²², reveals that among SNPs significantly associated with *IRX3* expression in brain or adipose tissue, only those associated with *IRX3* in brain show highly significant associations with BMI (Fig. 2c). Taken together, our data directly tie the noncoding genetic variation within *FTO* to tissue-specific modulation of *IRX3*, but not *FTO*, expression in human brain.

Next, we exploited animal models to determine a potential role for *IRX3* expression in the regulation of BMI and/or metabolism. Mice homozygous for an *Irx3*-null allele (*Irx3*-knockout mice) are viable and fertile,

with no evidence of embryonic lethality. We observed a 25–30% reduction of body weight in *Irx3*-knockout mice compared to control littermates (wild type), independent of gender (Fig. 3a and Extended Data Figs 6a, b and 7a, b). This difference becomes more pronounced if animals were subjected to a high-fat diet (HFD), with *Irx3*-knockout animals showing no significant body weight gain, contrasting to a 63% increase in control animals (Fig. 3a, b). Importantly, the percentage of fat mass in *Irx3*-knockout mice was significantly reduced without marked change of the lean mass ratio (Fig. 3c, d and Extended Data Fig. 7c, d). *Irx3*-knockout mice exhibit marked reduction in adiposity with smaller fat depots as well as reduced adipocyte size (Fig. 3e and Extended Data Figs 6b–d and 7e, f). These results were confirmed by differential gene expression of adiposity markers (that is, leptin, adiponectin and *mcp1*) in the perigonadal white adipose tissue (PWAT) of *Irx3*-knockout mice (Extended Data Fig. 6e). Importantly, *Fto* expression was not altered in the hypothalamus or PWAT of *Irx3*-knockout mice, suggesting that the lean phenotype of *Irx3*-knockout mice is not associated with *Fto* (Fig. 3f). As *Irx3*-knockout mice were resistant to HFD-induced obesity and metabolic disorder, such as hepatosteatosis (Fig. 3e), we examined glucose homeostasis by performing glucose tolerance tests (GTT) and insulin tolerance test (ITT) at different time points in the presence or absence of HFD. In 8-week-old mice, no difference in GTT was found (Fig. 3g). Although ageing and HFD led to glucose intolerance and insulin resistance in wild-type mice, *Irx3*-knockout mice showed none of these metabolic phenotypes (Fig. 3g–j).

Indirect calorimetric analysis showed higher energy expenditure in *Irx3*-knockout mice (Fig. 3k and Extended Data Fig. 8a, b, e). Importantly,

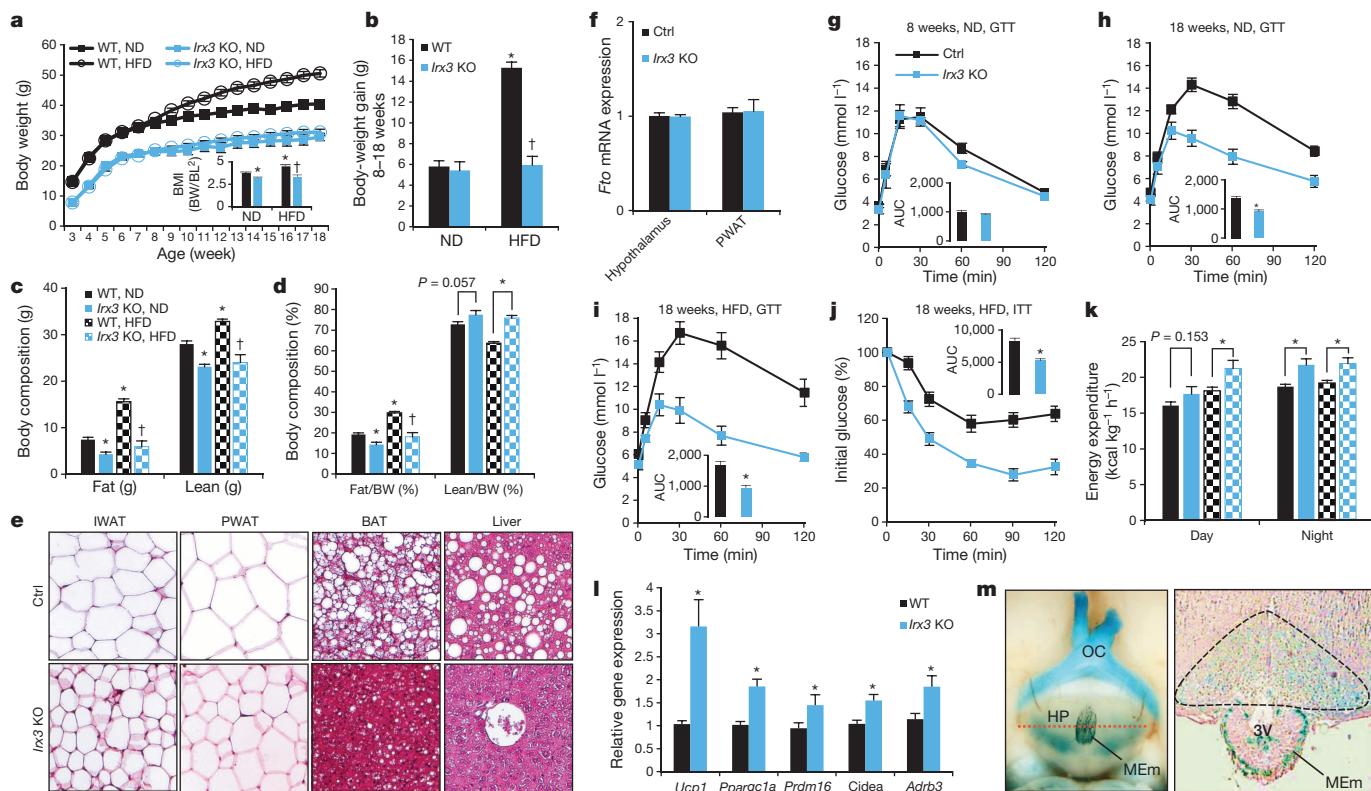


Figure 3 | *Irx3*-deficient mice are leaner and are protected against diet-induced obesity. **a**, Body weight in wild-type (WT) and *Irx3*-knockout (*Irx3* KO) mice fed a normal diet (ND) or high-fat diet (HFD). BL, body length. **b**, Weight gain in ND or HFD. **c**, Fat and lean mass in WT and *Irx3* KO mice. **d**, Fat and lean mass ratio as a percentage of body weight. **e**, Sections of inguinal white adipose tissue (IWAT; subcutaneous), perigonadal WAT (PWAT; visceral), brown adipose tissue (BAT), and liver from HFD mice. **f**, *Fto* messenger RNA expression in hypothalamus and PWAT of *Irx3* KO and WT mice. **g–i**, Glucose tolerance tests (GTT) in WT and *Irx3* KO mice. Inset graphs show area under curve (AUC). **j**, Insulin tolerance test (ITT) in HFD mice.

k, Energy expenditure of ND and HFD mice. **l**, Gene expression in PWAT. **m**, *Irx3* expression in the arcuate nucleus and median eminence (*Irx3*^{LacZ/+}). Left panel shows a ventral view of a whole-mount stained brain. A dashed red line indicates the position of cross-section, displayed in the right panel. β-galactosidase-stained area in arcuate nucleus is marked by a dashed black line. 3V, third ventricle; HP, hypothalamus; MEm, median eminence; OC, optic chiasm. Data are expressed as mean ± s.e.m. *P < 0.05 compared to WT or control; †P < 0.05 compared to each corresponding control group (see Methods). See additional statistical detail in Methods.

Irx3-knockout mice show upregulation of brown adipocyte markers in PWAT, including *Ucp1*, *Ppargc1a*, *Prdm16* and *cidea*, as well as increased expression of *Adrb3* encoding the β 3-adrenergic receptor, suggestive of elevated sympathetic activation (Fig. 3l). Furthermore, we found a significant increase of *Ucp1* expression in the brown adipose tissue (BAT) of *Irx3*-knockout mice (Extended Data Fig. 8e, f). WAT predominantly stores excessive calories and its accumulation causes obesity, whereas brown adipose tissue dissipates energy (lipids and glucose) and has been the focus for potential development of novel therapeutic strategies to treat obesity and diabetes. Our data suggest that ‘browning’ of WAT by higher sympathetic tone and activation of BAT might lead to increased energy expenditure in *Irx3*-knockout mice and partially account for their leanness and protection from diet-induced obesity.

Browning of WAT by increased sympathetic activity is a phenomenon controlled by hypothalamic circuits integrating central and peripheral metabolism regulation. Our human eQTL data suggest that the obesity-associated SNPs are associated with *IRX3* expression in brain. We determined that *Irx3* is expressed in the arcuate nucleus and median eminence of the hypothalamus, two critical regions involved in the regulation of energy homeostasis (Fig. 3m)^{23,24}. To investigate a possible role of *Irx3* expression in the hypothalamus mediating the body composition and energy homeostasis phenotypes in *Irx3*-knockout mice, we used *Rosa26^{EnR-Irx3}* conditional transgenic mice expressing a dominant negative form of *Irx3* (*EnR-Irx3*) by crossing with *Ins2-Cre* (Extended Data Fig. 9a, b)^{23,25}. This design allows for disrupting *Irx3* function while preserving the architecture of genomic interactions between the *Irx3* promoter and long-range regulatory sequences including the *Fto* obesity-associated region. Interestingly, *EnR-Irx3;Ins2-Cre*

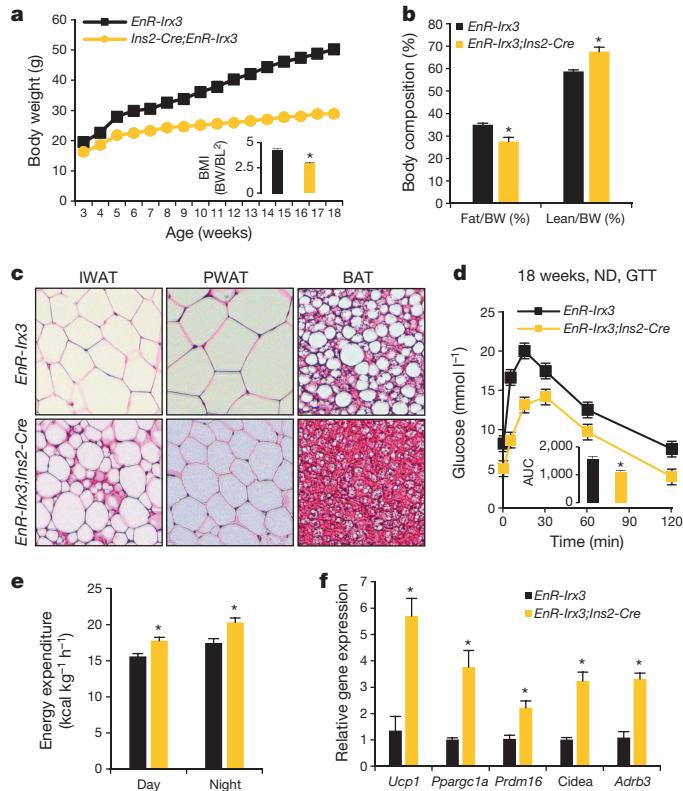


Figure 4 | Hypothalamus-specific dominant-negative *Irx3* mice recapitulate the metabolic phenotype of *Irx3*-deficient mice. **a**, Body weight of control (*EnR-Irx3*) and mutant mice (*EnR-Irx3;Ins2-Cre*) fed a normal diet. **b**, Fat and lean mass ratio as a percentage of body weight. **c**, Haematoxylin & eosin (H&E)-stained sections of IWAT, PWAT and BAT. **d**, Glucose tolerance test. **e**, Energy expenditure corrected for lean mass. **f**, Quantitative PCR in PWAT in control and mutant mice. Data are expressed as mean \pm s.e.m. * $P < 0.05$ compared to control group. Additional details are given in Methods.

mice phenocopy the physiological, histological, and molecular metabolism phenotypes of germline *Irx3*-knockout mice (Fig. 4a–d and Extended Data Fig. 9c, d). Energy expenditure was significantly higher in *EnR-Irx3;Ins2-Cre* mice with similar food intake and locomotor activity (Fig. 4e and Extended Data Fig. 10a–e). Notably, these mice also showed ‘browning’ of WAT as well as activation of BAT (Fig. 4f and Extended Data Fig. 10f, g). Together, our data support the notion that the hypothalamic expression of *Irx3* regulates energy homeostasis and body composition.

Our results do not dispute the idea emerging from animal models that *Fto* is also a regulator of body mass and composition. The functional link between *Fto* and body mass regulation originated from animal models, unequivocally demonstrating that variation in *Fto* expression results in variation in body mass phenotypes. But to assess the null expectation that genetic manipulations in mice may result in variation in these parameters, we searched among 7,556 targeted mouse gene knockout models for evidence of alterations in body size, mass and growth (see Methods). Nearly one-third (2,166; 29%) of gene knockouts in mice show alteration of these phenotypes, underscoring that animal models alone are not sufficient to definitively establish a functional relationship between a given gene and long-range noncoding variants (Supplementary Table 4). It is the aggregate of our data on chromatin looping, patterns of *in vivo* enhancer function, human eQTL mapping and mouse model data that collectively place *IRX3* as an important target gene of the genetic association with human obesity. These observations have direct implications for how to interpret the associations characterized for thousands of noncoding variants with human disease and related quantitative traits. We observe more than 150 genomic regulatory blocks, similar to that containing *Fto* and *IRX3*, overlapping with noncoding SNPs associated with human disease and complex traits in GWAS (Methods and Supplementary Table 5), highlighting the need for a careful experimental pipeline to define the target gene(s) for each disease-associated noncoding SNP.

Our data posit that the obesity-associated SNPs within *Fto* are functionally connected with regulation of *IRX3* expression, and that *IRX3* is an important determinant of body mass and composition. *IRX3* encodes a transcription factor highly expressed in brain. Although our data represent the first demonstration of the intersection of *IRX3* biology with body mass composition and metabolism, previous work identified *IRX3* overexpression in adipocytes as a hallmark of the molecular switch seen in patients after profound weight loss following bariatric surgery²⁶. These data indicate that *IRX3* may have important roles regulating metabolism beyond the ones we describe associated with *Irx3* expression in hypothalamus. Future investigations will determine the precise molecular mechanisms by which *IRX3* regulates metabolic parameters.

METHODS SUMMARY

For 4C-seq, chromatin was digested with DpnII and Csp6I. Captured DNA was amplified using promoter-specific primers and deep sequenced.

For 3C, nuclei were digested with HindIII. Primer quality was assessed using serial dilutions of BACs encompassing the regions of interest (RP23-268O10, RP23-96F3). The average of four independent experiments is represented graphically (Extended Data Fig. 2c).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 August 2013; accepted 10 February 2014.

Published online 12 March 2014.

- Dina, C. et al. Variation in *Fto* contributes to childhood obesity and severe adult obesity. *Nature Genet.* **39**, 724–726 (2007).
- Frayling, T. M. et al. A common variant in the *Fto* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
- Scuteri, A. et al. Genome-wide association scan shows genetic variants in the *Fto* gene are associated with obesity-related traits. *PLoS Genet.* **3**, e115 (2007).
- Church, C. et al. Overexpression of *Fto* leads to increased food intake and results in obesity. *Nature Genet.* **42**, 1086–1092 (2010).

5. Fischer, J. et al. Inactivation of the *Fto* gene protects from obesity. *Nature* **458**, 894–898 (2009).
6. Gao, X. et al. The fat mass and obesity associated gene *FTO* functions in the brain to regulate postnatal growth in mice. *PLoS ONE* **5**, e14005 (2010).
7. Grunnet, L. G. et al. Regulation and function of *FTO* mRNA expression in human skeletal muscle and subcutaneous adipose tissue. *Diabetes* **58**, 2402–2408 (2009).
8. Klöting, N. et al. Inverse relationship between obesity and *FTO* gene expression in visceral adipose tissue in humans. *Diabetologia* **51**, 641–647 (2008).
9. Wählén, K., Sjolin, E. & Hoffstedt, J. The common rs9939609 gene variant of the fat mass- and obesity-associated gene *FTO* is related to fat cell lipolysis. *J. Lipid Res.* **49**, 607–611 (2008).
10. McMurray, F. et al. Adult onset global loss of the *fto* gene alters body composition and metabolism in the mouse. *PLoS Genet.* **9**, e1003166 (2013).
11. Jin, F. et al. A high-resolution map of the three-dimensional chromatin interactions in human cells. *Nature* **503**, 290–294 (2013).
12. Houweling, A. C. et al. Gene and cluster-specific expression of the *Iroquois* family members during mouse development. *Mech. Dev.* **107**, 169–174 (2001).
13. van Tuyl, M. et al. *Iroquois* genes influence proximo-distal morphogenesis during rat lung development. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **290**, L777–L789 (2006).
14. Gerken, T. et al. The obesity-associated *FTO* gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469–1472 (2007).
15. Qi, L. et al. Fat mass-and obesity-associated (*FTO*) gene variant is associated with obesity: longitudinal analyses in two cohort studies and functional test. *Diabetes* **57**, 3145–3151 (2008).
16. Stratigopoulos, G. et al. Regulation of *Fto/Ftm* gene expression in mice and humans. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **294**, R1185–R1196 (2008).
17. Ragvin, A. et al. Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to *HHEX*, *SOX4*, and *IRX3*. *Proc. Natl Acad. Sci. USA* **107**, 775–780 (2010).
18. Visel, A. et al. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35** (Database issue), D88–D92 (2007).
19. Bosse, A. et al. Identification of the vertebrate *Iroquois* homeobox gene family with overlapping expression during early development of the nervous system. *Mech. Dev.* **69**, 169–181 (1997).
20. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
21. Gamazon, E. R. et al. Enrichment of *cis*-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol. Psychiatry* **18**, 340–346 (2012).
22. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42**, 937–948 (2010).
23. Kong, D. et al. GABAergic RIP-Cre neurons in the arcuate nucleus selectively regulate energy expenditure. *Cell* **151**, 645–657 (2012).
24. Shi, Y. C. et al. Arcuate NPY controls sympathetic output and BAT function via a relay of tyrosine hydroxylase neurons in the PVN. *Cell Metab.* **17**, 236–248 (2013).
25. Mori, H. et al. Critical role for hypothalamic mTOR activity in energy balance. *Cell Metab.* **9**, 362–374 (2009).
26. Dankel, S. N. et al. Switch from stress response to homeobox transcription factors in adipose tissue after profound fat loss. *PLoS ONE* **5**, e11033 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank F. Gage, C. Marchetto, B. Ren and F. Jin for their generosity in sharing reagents and data. This work was funded by grants from the National Institutes of Health (DK093972, HL119967, HL114010 and DK020595) to M.A.N. and (MH101820, MH090937 and DK20595) to N.J.C. J.L.G.-S. was funded by grants from the Spanish Ministerio de Economía y Competitividad (BFU2010-14839, CSD2007-00008) and the Andalusian Government (CVI-3488). C.-C.H. was supported by a grant from the Canadian Institute of Health Research. K.-H.K. is supported by a fellowship from the Heart and Stroke Foundation of Canada. S.S. is supported by an NIH postdoctoral training grant (T32HL007381).

Author Contributions M.A.N., J.L.G.-S. and C.-C.H. designed the project. J.J.T. and C.G.-M. performed 4C-seq experiments, with analysis also aided by S.S. and N.J.S. N.J.S. performed the locus conservation and regulatory block analysis. I.A., F.L.C., D.R.S., N.F.W. and S.S. performed *in vivo* enhancer experiments. S.S. and M.S. performed *in situ* hybridizations and mouse knockout phenotype calculation. E.R.G. and N.J.C. performed eQTL analyses. K.-H.K. performed mouse metabolic experiments. J.H.L. and D.T. contributed to histological analysis and glucose homeostasis analysis. V.P. contributed to gene and protein expression analysis. J.E.S. contributed to metabolic cage analysis. H.K.S., D.R.S., M.M., S.N., N.A.V., R.D.A. and A.N. provided scientific discussion and technical support. S.S., C.-C.H., J.L.G.-S. and M.A.N. wrote the paper with input from all authors.

Author Information Data were submitted to GEO under accession GSE52830. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.N. (nobrega@uchicago.edu), J.L.G.-S. (jlgomska@upo.es) or C.-C.H. (cchui@sickkids.ca).

METHODS

3C assays. 3C assays were performed as described in ref. 27. Mouse brain tissue was processed to get single cells samples. Ten-million isolated cells were treated with lysis buffer (10 mM Tris-HCl, pH 8, 10 mM NaCl, 0.3% IGEPAL CA-630 (Sigma-Aldrich, I8896), 1X protease inhibitor cocktail (cComplete, Roche, 11697498001)) and nuclei were digested with HindIII endonuclease (Roche, 10798983001). Subsequently, DNA was ligated with T4 DNA ligase (Promega, M1804). A set of locus specific primers (see table) was designed with the online program primer3 v. 0.4.0 (ref. 28), each one close to a HindIII restriction site. These primers were used to make real-time quantitative PCRs using a BioRad CFX96 Real Time System with SsoFast EvaGreen Supermix (BioRad 172-5203), in order to measure the relative enrichment in each ligation product. The primer designed within the linkage disequilibrium-block was used as fixed primer. The quality of all primer pairs was measured using serial dilutions of a bacterial artificial chromosome (BAC) DNA that encompass the regions of interest (RP23-268O10, RP23-96F3). PCR values were normalized by means of control primers designed in the *Erc3* gene locus. The average of four independent experiments is represented graphically (Extended Data Fig. 2c).

4C-seq. 4C-seq assays were performed as previously reported^{27,29–31}. Whole embryos (9.5 days of development) or whole adult (8 weeks) mice brains (both strain CD-1; Charles River) were processed to get approximately 10-million isolated cells, which were treated with lysis buffer (10 mM Tris-HCl pH 8, 10 mM NaCl, 0.3% IGEPAL CA-630 (Sigma-Aldrich, I8896), 1X protease inhibitor cocktail (cComplete, Roche, 11697498001)). Nuclei were digested with DpnII endonuclease (New England Biolabs, R0543M) and ligated with T4 DNA ligase (Promega, M1804). Subsequently, Csp6I endonuclease (Fermentas, Thermo Scientific, FD0214) was used in a second round of digestion, and the DNA was ligated again. Specific primers were designed near each gene promoter with primer3. Viewpoint fragend coordinates (for mouse genome build m9) were as follows: *Fto*, chromosome 8, 93,837,310–93,837,724; *Irx3*, chromosome 8, 94,325,277–94,326,009. Illumina adaptors were included in the primers sequence³². Sixteen PCRs were performed with Expand Long Template PCR System (Roche, 11759060001) for each viewpoint, pooled together and purified using High Pure PCR Product Purification Kit (Roche, 11732668001). Quanti-T_iT PicoGreen dsDNA Assay Kit (Invitrogen, P11496) was used in order to measure sample concentration, and then deep sequenced. 4C-seq data were analysed, with some changes, as described in ref. 30.

In brief, raw sequencing data were de-multiplexed and aligned using either the Mouse July 2007 (NCBI37/mm9) or Zebrafish December 2008 (Zv8/danRer6) assembly as the reference genome. Reads located in fragments flanked by two restriction sites of the same enzyme, or in fragments smaller than 40 bp were filtered out. Mapped reads were then converted to reads-per-1kb-bin units, and smoothed using a running mean 5-fragment window algorithm. To calculate the statistically significant targets for each viewpoint, a background theoretical model was calculated as the exponential fit curve of the average signal of 61 4C-seq different samples^{11,33,34}. The *P* value for each bin was calculated by means of Poisson probability function. Smoothed data were uploaded to the UCSC Browser³⁵ for visualization and plotted using the browser and Circos³⁶. For plotting, statistically significant targets were defined as *P* < 0.01.

Primers used for 3C experiments. The following primers were used for 3C experiments: '3c_mouse_LDBlock', 5'-TGGTCTCGGGTATCTGTCC-3'; '3c_mouse_Irx3_control', 5'-TCGCTCTAGTGAGACTTTC-3'; '3c_mouse_Irx3_promoter', 5'-TGATGTTGGTCCCTACTAGG-3'; '3c_mouse_Irx3_control2', 5'-TCCACAGAAACATCTGACG-3'; '3c_mouse_Erc3-1', 5'-TGACCTCCACACTCCTGAC-3'; '3c_mouse_Erc3-2', 5'-ATGCGAATTAGAAACTGC-3'.

Primers used for 4C-seq experiments. The following primers were used for 4C-seq experiments: '4c_mouse_Fto_read', 5'-AATGATACGGCGACCACCGAAC ACTCTTCCCTACACGACGCTTCC GATCTCATATTGCTCTGGATGCA GATC-3'; '4c_mouse_Fto_noread', 5'-CAAGCAGAACGGCATACGAGAA CTACCCCTCCCAGAATGC-3'; '4c_mouse_Irx3_read', 5'-AATGATACGGCG ACCACCGAACACTCTTCCCTACACGACGCTCTCCGAT CTCCGCCGC GGAGCAGATC-3'; '4c_mouse_Irx3_noread', 5'-CAAGCAGAACGGCATACGATGGAGTCGCAATCACC-3'; '4c_zebrafish_Fto_read', 5'-AATGATAC GGGCACCACCGAACACTCTTCCCTACACGACGCTCTCCGAT CTCCCT CCACTGTCATCCGATC-3'; '4c_zebrafish_Fto_noread', 5'-CAAGCAGAAG CGGCATACGACTACATTATGTCAGTTGGG-3'; '4c_zebrafish_Irx3a_read', 5'-AATGATACGGCGACCACCGAACACTCTTCCCTACACGACGCTCTT CCGAT CTCCCTACCGGATTACTCTACAGATC-3'; '4c_zebrafish_Irx3a_noread', 5'-CAAGCAGAACGGCATACGAAAAACGCCAGAAGACTGAC-3'.

ENCODE ChIA-PET. Publicly available data for the breast cancer MCF7 cell line, generated as part of the ENCODE project³⁷, were downloaded from and visualized with the WashU EpiGenome Browser³⁸ (<http://epigenomegateway.wustl.edu/browser/>). ChIA-PET has been performed to capture long-range chromatin interactions associated with RNA polymerase II³⁹.

Locus conservation. To assess the level of high-level conservation of the *FTO-IRX5* locus, we used phastCons⁴⁰ elements and scores as a measurement of conservation.

High-scoring phastCons elements (LOD > 680, top 1%) cover 3% (38,040 out of 1,244,409 bp) of the genomic space that includes *FTO* and *IRX5* (hg19, chromosome 16: 53,731,249–54,975,288), whereas on average, only $0.5 \pm 0.6\%$ of genomic intervals with the same length are occupied by such high-scoring phastCons elements. Genomic intervals of the same size of the *FTO-IRX3* locus were generated every 10 kb for each chromosome. The *FTO-IRX5* is at the top 1% intervals containing high-scoring phastCons elements.

Similarly, in the *FTO-IRX5* interval, 9% of the base pairs have phastCons score above 0.9 (range: 0.0–1.0), whereas 14,164 intervals from all chromosomes of the same size offset by 200 kb contain only $3.8\% \pm 1.9\%$ of such high-scoring base pairs. At 9%, the *FTO-IRX5* is at the top 2% intervals with the most conserved base pairs.

These results indicate that the *FTO-IRX5* interval contains significantly more highly conserved base pairs than most genomic loci.

Mouse *in vivo* transgenic reporter assays. Enhancer reporter assays and constructs, including the modified BAC, were created as reported previously⁴¹. The human (NCBI37/hg19) genomic coordinates of the elements cloned are: *FTO*-1, chromosome 16, 53,800,270–53,805,184; *FTO*-2, chromosome 16, 53,808,432–53,813,121; *FTO*-3, chromosome 16, 53,816,312–53,821,148. *FTO*-promoter, chromosome 16, 53,736,860–53,738,096; *IRX3*-promoter, chromosome 16, 54,319,963–54,322,782. BAC RP11-261B9 was modified by inserting a *LacZ*-Amp cassette in place of the 36 nucleotides in exon 1 at chromosome 16, 53,738,097–53,738,132. This homologous recombination was facilitated by using the following homology arms: *FTO*-hArm1, chromosome 16, 53,738,047–53,738,096; *FTO*-hArm2, chromosome 16, 53,738,133–53,738,182.

Mouse *in vivo* transgenic reporter assays were performed as reported previously⁴¹. *In situ* hybridization assays were performed according to standard protocols⁴². The full-length mouse *Fto* coding DNA sequence (CDS) as riboprobe template was from IMAGE clone 5708558 (ATCC).

Gene expression in human tissues. For the eQTL analyses, we used cerebellum²¹ (*n* = 153; publicly available in GEO under accession number GSE35974) and adipose⁴³ (*n* = 62; publicly available in GEO under accession number GSE40234); details of the samples, the quality control procedure and the normalization approach used were described previously²¹. For adipose, the samples were chosen to be at the tails of the distribution of insulin sensitivity matched on age, gender and natural log of BMI. Because the adipose samples included individuals of African ancestry, we performed principal component analysis and tested each marker in the locus for the additive effect of allele dosage on the residual expression phenotype (IRX3 and FTO) after adjusting for the principal components (*n* = 2). Similarly, for cerebellum, we performed linear regression on the residual expression phenotype (IRX3 and FTO) after adjusting for sex and pH to evaluate the additive effect of allele dosage at each marker in the locus; the non-diseased samples (obtained from the Stanley Medical Research Institute) included in the analysis were of European descent and thus the principal components were not used to generate the residual expression. Significance (*P* value) was evaluated using the *t*-statistic from the regression.

Single nucleotide polymorphisms associated with gene expression levels. We downloaded genome-wide association data from the Genetic Investigation of Anthropometric Traits²² (GIANT) Consortium website (http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consorium_data_files#BMI_28down_load_GZIP.29). We extracted the *P* values for those SNPs in the locus with significant association (*P* < 0.05) with IRX3 or FTO expression in adipose or cerebellum. We plotted the distribution of association *P* values for these expression-associated SNPs (eSNPs). rs9930506 is highly associated with BMI (*P* = 1.41×10^{-53} , *n* = 123541). Recombination rates are estimated from the International HapMap Project.

Statistical analyses. Statistical analyses were carried out using the R statistical software (<http://cran.r-project.org/>). Regional plots were done by repurposing an R script retrieved from the Broad Institute (<http://www.broadinstitute.org/diabetes/scandinavs/figures.html>).

Mice. All animal experimental protocols approved by the Animal Care Committee of the Toronto Centre of Phenogenomics conformed to the standards of the Canadian Council on Animal Care. *Irx3*-deficient (*Irx3*-knockout) mice⁴⁴ and *Ins2-Cre* mice²³ were described previously. *Rosa26^{EnR-Irx3}* conditional transgenic mice (*EnR-Irx3*) were generated using pROSA26PA gene-targeting vector as described previously⁴⁵. Mice were maintained on 12-h light-dark cycles and provided with food and water *ad libitum*. For diet-induced obesity studies, 8-week-old male mice were subjected to 45% high-fat diet (Research Diets) for 10 weeks. Body weight was measured every week from 4 to 18 weeks of age, and body composition was analysed using an EchoMRI device (Echo Medical Systems) at 18 weeks of age. Body mass index (BMI) was calculated by dividing body weight (g) by body length (mm) squared (BMI = body weight/body length²).

Metabolic phenotyping experiments. Energy expenditure was evaluated through indirect calorimetry (Oxymax System, Columbus Instruments) over periods of 24 h.

In brief, energy expenditure was calculated by multiplying oxygen consumption (VO_2) by the calorific value ($\text{CV} = 3.815 + 1.232 \times \text{respiratory exchange ratio}$) and corrected by lean mass. Locomotor activity and food intake were also measured simultaneously. For glucose and insulin tolerance tests, mice were subjected to intraperitoneal bolus injection of glucose (1 mg g⁻¹ of body weight) or insulin (1.5 milli-units per g) after fasting for overnight (14 to 16 h) or 6 h with water *ad libitum*, respectively. Blood glucose levels were measured at the indicated intervals. For histological analysis, tissues were fixed in 4% paraformaldehyde and embedded in paraffin. Sections of 5 μm were stained with haematoxylin and eosin (H&E). *Irx3* expression in brain and lung was examined using *Irx3*^{lacZ} knock-in mice with β -galactosidase staining.

Gene and protein expression analyses. Total RNA was extracted from fat tissues or hypothalamus using RNeasy Lipid Tissue Kit (Qiagen), and complementary DNA was synthesized from 2 μg of RNA using M-MLV reverse transcriptase (Invitrogen) with oligo(dT). Gene expression assay was conducted using SYBR Green methods on ViiA7 (Applied Biosystems), and relative cycle threshold (CT) values were normalized by β -actin. Primer sequences are available upon request. Western blot analysis was performed as described previously⁴⁶. The same amount of protein extracted from hypothalamus or brown adipose tissues was loaded, which was confirmed by β -actin as a loading control. Antibodies used are as follows: Ucp1 (Santa Cruz Biotechnology), *Irx3* (generated in-house) and β -actin (Calbiochem).

Statistics. All results are expressed as mean \pm s.e.m. Statistical significance of differences among groups was determined by Student's *t*-test or analysis of variance (ANOVA) with post-hoc analysis, Student–Newman–Keuls using Sigma Stat (SPSS). Differences at $P < 0.05$ were considered statistically significant (* $P < 0.05$ versus WT or control; † $P < 0.05$ versus WT, HFD). The following numbers of mice per sample were used for comparison; body-weight measurement for *Irx3* mice (ND, WT/knockout (KO), $n = 32/25$; HFD, WT/KO, $n = 12/7$) and *EnR-Irx3* mice (control/mutant, $n = 9/12$); body composition analysis for *Irx3* mice (ND WT/KO, $n = 21/12$; HFD WT/KO, $n = 8/5$) and *EnR-Irx3* mice (control/mutant, $n = 9/12$); glucose tolerance test for *Irx3* mice (ND 8 weeks, WT/KO, $n = 8/5$; ND 18 weeks, WT/KO, $n = 10/6$; HFD 18 weeks, WT/KO, $n = 12/7$) and *EnR-Irx3* mice (control/mutant, $n = 5/8$); insulin tolerance test for *Irx3* mice (WT/KO, $n = 12/7$); indirect calorimeter analysis for *Irx3* mice (ND WT/KO, $n = 7/5$; HFD WT/KO, $n = 8/4$) and *EnR-Irx3* mice (control/mutant, $n = 5/7$), and gene expression analysis for *Irx3* mice (WT/KO, $n = 10/7$) and *EnR-Irx3* mice (control/mutant, $n = 5/7$).

Mouse body-mass phenotypes. To calculate the fraction of knockout mice that display a phenotype affecting body mass, we consulted the Mouse Genome Informatics (MGI) database⁴⁷.

First, to determine the number of gene knockouts that result in a body mass phenotype, we needed to define a more precise phenotype. The Mammalian Phenotype Ontology⁴⁸ (http://www.informatics.jax.org/searches/MP_form.shtml) specifies a controlled vocabulary by which all mice in the MGI database are classified. Within this framework, (we believe) the most applicable is 'abnormal postnatal growth/weight/body size' (MP:0002089).

Next, we crafted a query using the MouseMine portal⁴⁹ (<http://www.mousemine.org>) to access the MGI database, using MP:0002089 as the search criteria in the 'Mammalian phenotypes → Mouse genotypes' template (http://www.mousemine.org/mousemine/template.do?name=MPhenotype_MFeature&scope=all). We asked for all targeted alleles on chromosomes 1–19, X and Y, reasoning that many non-targeted (that is, random, chemical and radiation-induced) mutations have not been precisely mapped, making it difficult if not impossible to ensure that they are not counted multiple times. Also, in the context of our study, we are discussing the chance that a targeted knockout would give a specific phenotype, so these types of alleles are most relevant. Finally, we excluded alleles present only as cell lines, which left 2,166 unique genes.

To determine the total number of targeted mutations present in the MGI database, we performed a search identical to that above but omitting the MP:0002089 term, which left 7,556 unique genes.

Genomic regulatory blocks. Genomic regulatory blocks are regions under strong selection that contain a number of syntenic highly conserved sequences, believed to regulate a syntenic target gene ('anchor gene', the gene that is the *raison d'être* of the GRB). There are different approaches to determine GRBs, usually identifying syntenic genes and non-coding conserved regions that span a region and determining the boundaries of the region.

GWAS identified SNPs that might affect noncoding regulatory elements that, in turn, regulate specific target genes. Assignment of regulatory elements to targets is

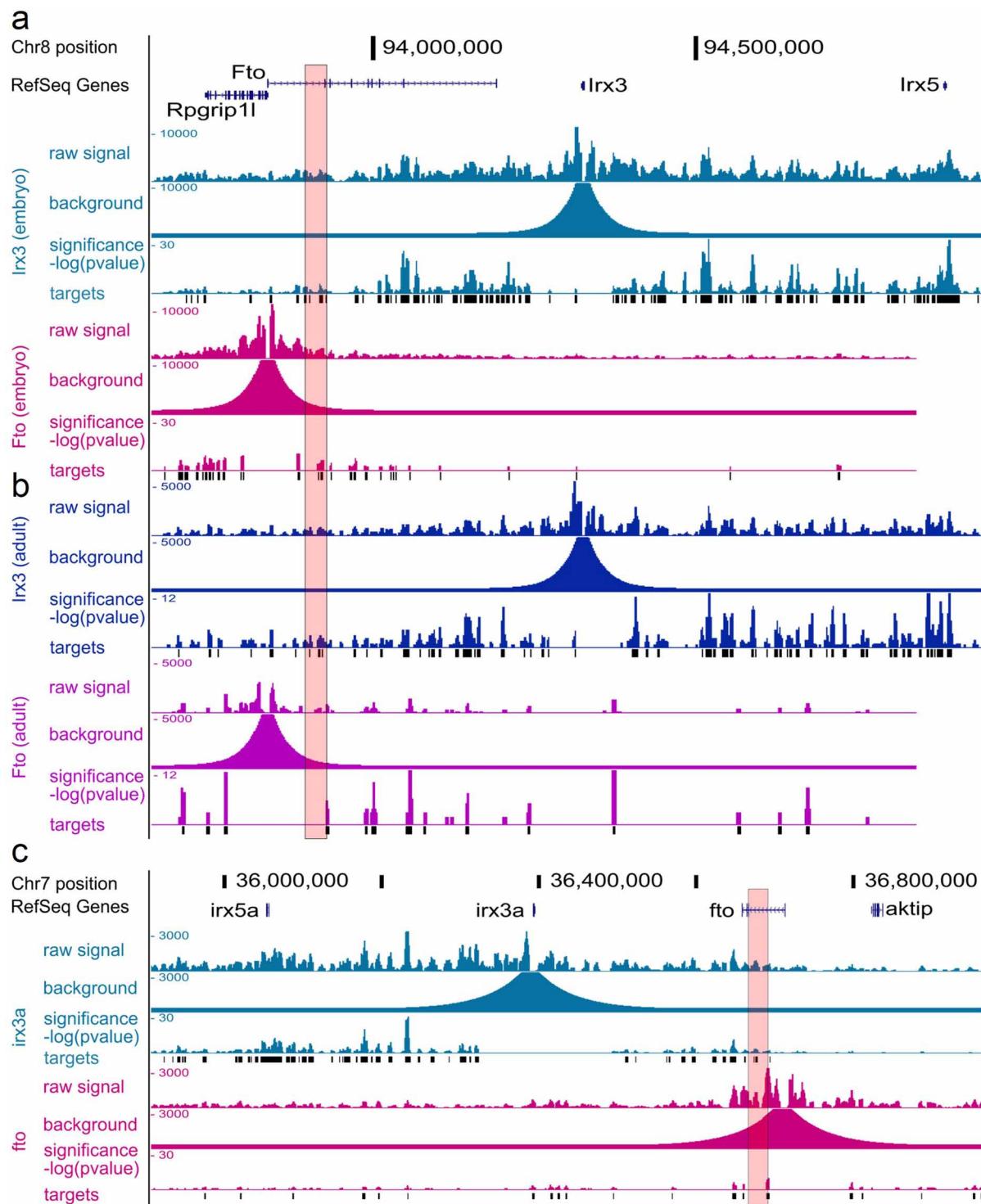
still a challenge, the simplest method being to assign them to the nearest gene. In the case of FTO, because the GWAS SNP occurs in an FTO intron, this gene was considered to be the target.

The idea put forward here is that when regulatory elements occur within a GRB, which is the case for FTO, IRX3 and IRX5, it is likely that they target the anchor gene (*IRX3* or *IRX5*), given its importance. Supplementary Table 5 contains an annotation of GRBs and their anchor genes, and the GWAS SNPs within the GRB, as well as annotation of whether the nearest gene is the anchor gene or not. In those cases in which the nearest gene differs from the anchor gene, we propose that the anchor gene be used as the putative target and additional experimentation as shown in this work should be performed to determine precisely which is the target gene for the GWAS-affected regulatory element.

Genomic regulatory blocks were obtained from UCNEbase (<http://ccg.vital-it.ch/UCNEbase/>)⁵⁰. Human UCNE clusters (hg19) were downloaded from http://ccg.vital-it.ch/UCNEbase/data/download/clusters/hg19_clusters_coord.bed and are referred here as GRBs.

The hg19 GWAS catalogue from the UCSC Genome Browser was downloaded and the tagging SNPs contained in GRBs with more than one gene (making SNP assigning doubtful) were kept for further annotation. SNPs occurring in exons were removed. The nearest gene was considered to be the RefSeq gene whose transcription start site was the nearest, or the gene where the SNP is contained.

27. Hagège, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature Protocols* **2**, 1722–1733 (2007).
28. Rozen, S. & Skaltsky, H. Primer3 on the WWW for general users and for biologist programmers. *Meth. Mol. Biol.* **132**, 365–386 (2000).
29. Dekker, J. *et al.* Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
30. Noordermeer, D. *et al.* The dynamic architecture of *Hox* gene clusters. *Science* **334**, 222–225 (2011).
31. Splinter, E. *et al.* Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58**, 221–230 (2012).
32. Stadhouders, R. *et al.* Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nature Protocols* **8**, 509–524 (2013).
33. Denholtz, M. *et al.* Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13**, 602–616 (2013).
34. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
35. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
36. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
37. ENCODE. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
38. Zhou, X. & Wang, T. Using the Wash U Epigenome Browser to examine genome-wide sequencing data. *Curr. Protoc. Bioinformatics* **40**, 10.10.1–10.10.14 (2012).
39. Li, G. *et al.* Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
40. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–50 (2005).
41. Smemo, S. *et al.* Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* **21**, 3255–3263 (2012).
42. Wilkinson, D. G. & Nieto, M. A. Detection of messenger RNA by *in situ* hybridization to tissue sections and whole mounts. *Methods Enzymol.* **225**, 361–373 (1993).
43. Elbein, S. C. *et al.* Genetic risk factors for type 2 diabetes: a trans-regulatory genetic architecture? *Am. J. Hum. Genet.* **91**, 466–477 (2012).
44. Zhang, S. S. *et al.* Iroquois homeobox gene 3 establishes fast conduction in the cardiac His-Purkinje network. *Proc. Natl. Acad. Sci. USA* **108**, 13576–13581 (2011).
45. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001).
46. Li, Z. J. *et al.* Kif7 regulates Gli2 through Sufu-dependent and -independent functions during skin development and tumorigenesis. *Development* **139**, 4152–4161 (2012).
47. Eppig, J. T. *et al.* The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res.* **40** (Database issue), D881–D886 (2012).
48. Smith, C. L. & Eppig, J. T. The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Biol. Med.* **1**(3), 390–399 (2009).
49. Smith, R. N. *et al.* InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. *Bioinformatics* **28**, 3163–3165 (2012).
50. Dimitrieva, S. & Bucher, P. UCNEbase—a database of ultraconserved non-coding elements and genomic regulatory blocks *Nucleic Acids Res.* **41** (Database issue), D101–D109 (2013).

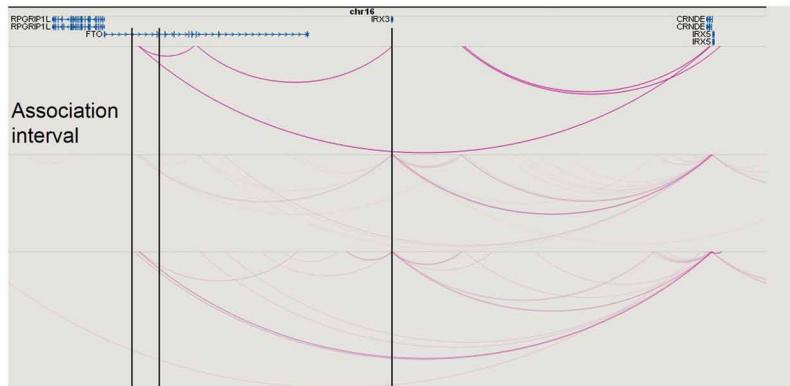
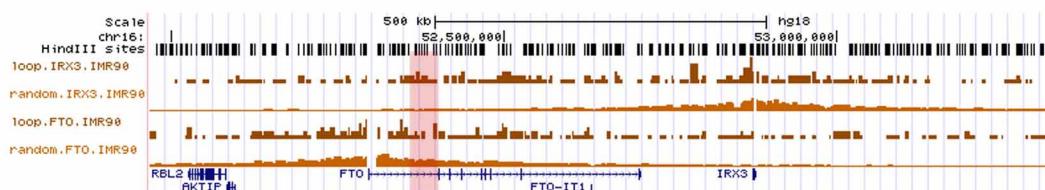
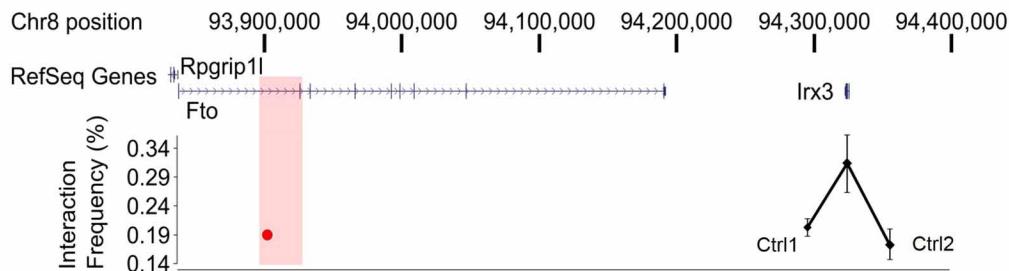


Extended Data Figure 1 | Long-range interactions in mouse and zebrafish. 4C-seq data for the *Fto*-*Irx3* locus, visualized with the UCSC Genome Browser. **a**, Data (also shown in the circular plot in Fig. 1) generated using whole mouse embryos (E9.5), showing the frequency of interactions with the promoter of *Irx3* (blue, top) or *Fto* (magenta, bottom). The background signal corrects for the strong correlation between (nonspecific) ligation events and the linear

distance along the chromosome. Poisson statistical significance ($-\log(P\text{ value})$) of the 4C-seq interactions over the background is plotted. Significant interactions ($P < 0.01$), ‘targets’, are displayed in black. **b**, As above for **a** but for adult mouse brain (8 weeks). **c**, As above for **a** but for whole zebrafish embryos (24 h post fertilization). In all, the region orthologous to the obesity association interval in the first intron of *Fto* is highlighted in pink.

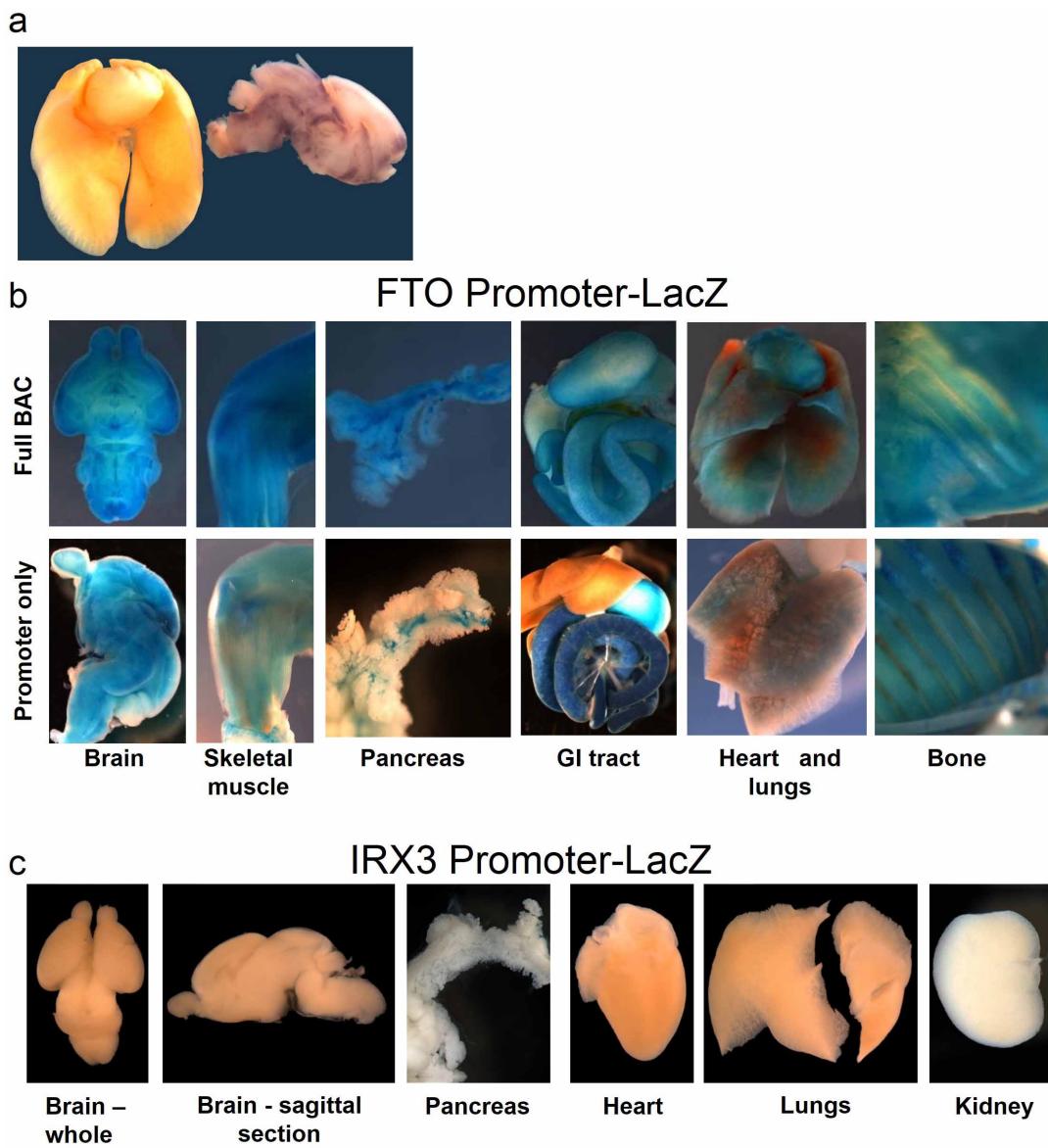
a

RefSeq Genes

gisChIA-PET MCF7
RNAPII pilot rep1ENCODE GIS ChIA-
PET MCF-7 POL2 rep3ENCODE GIS ChIA-
PET MCF-7 POL2 rep4**b****c**

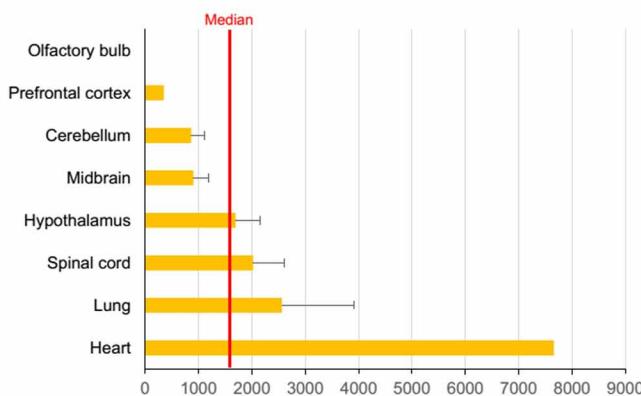
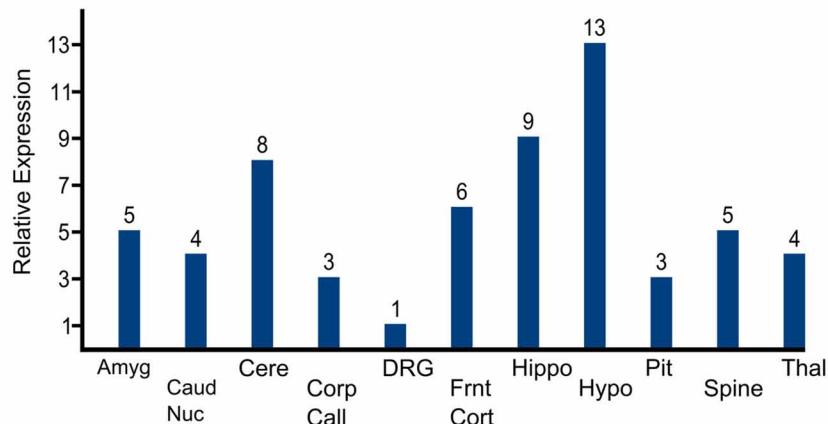
Extended Data Figure 2 | Long-range interactions at the *FTO*-*IRX3* locus. **a**, ENCODE data for ChIA-PET using RNA polymerase 2 (POL2) in MCF7 (human breast adenocarcinoma) cells shows interactions between *IRX3* and the obesity association interval in the first intron of *FTO*. No interactions are observed between the *FTO* promoter and the association interval. These public data are available from and was visualized with the WashU EpiGenome Browser (<http://epigenomegateway.wustl.edu/browser/>). **b**, Hi-C data

previously generated¹¹ in human IMR-90 (fetal lung) cells. In the association interval, the *IRX3* signal is stronger than the background (random) signal. However, the signal for *FTO* is not. **c**, 3C data generated with adult (8 weeks) mouse brain. Using bait (red circle) in the association interval (red rectangle), we observe more frequent interactions with the *Irx3* promoter compared to control regions 1 and 2 that are 29 and 42 kb away, respectively, indicative of looping.



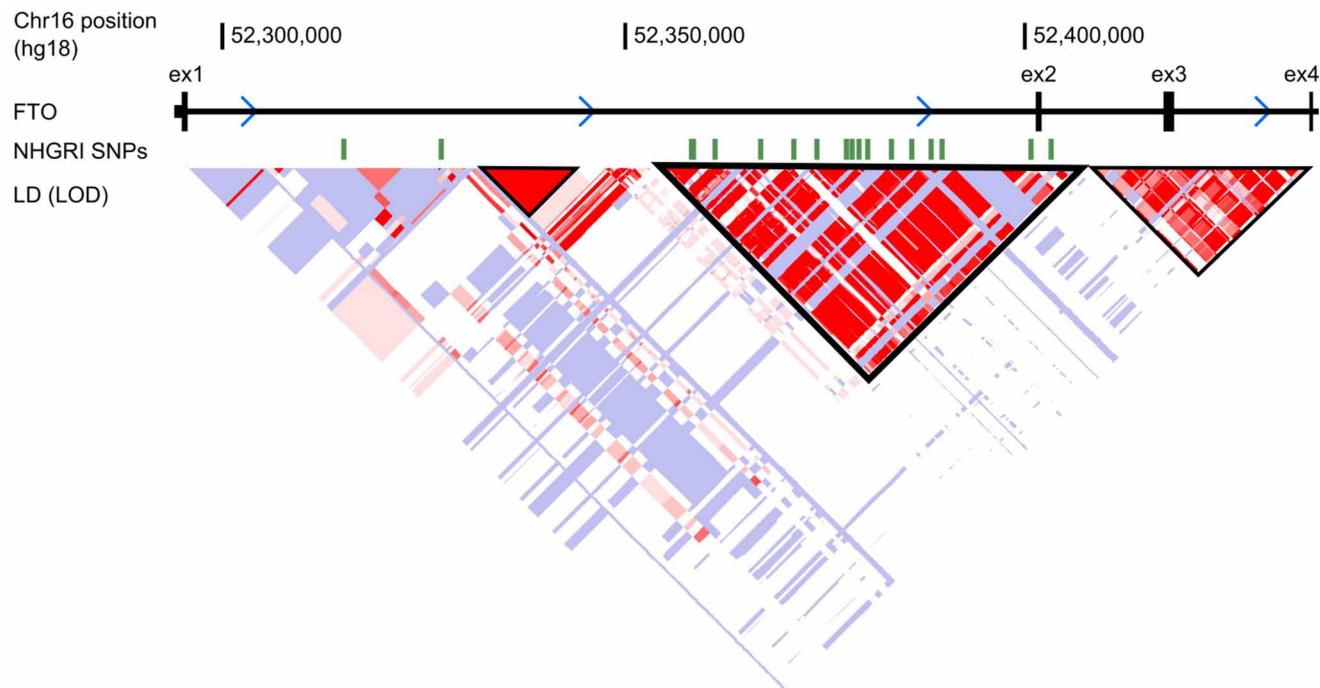
Extended Data Figure 3 | Gene expression in mouse tissue. **a,** *FTO* expression in lung and brain, shown by RNA *in situ* hybridization for mouse *Fto* mRNA, in newborn (P1) mouse. Lungs and heart (left, whole organs) were processed simultaneously and in the same well as brain (right, sagittal section) so that the relatively higher expression in brain can be observed. **b,** LacZ staining for β -galactosidase expression driven from the human *FTO* promoter. Top, the promoter-LacZ construct is isolated: only the 1,237 bp proximal to the transcriptional start site are included. Broad expression is recapitulated, indicating the robust transcriptional competency of the human *FTO* promoter. **c,** In contrast, the 2,820-bp proximal human *IRX3* promoter is not sufficient to drive LacZ expression, which is consistent with an enhancer-dependent transcriptional control mechanism.

obesity-associated interval and any enhancers present. The broad expression is consistent with previous reports in human and mouse (see main text for references). At bottom, the promoter-LacZ construct is isolated: only the 1,237 bp proximal to the transcriptional start site are included. Broad expression is recapitulated, indicating the robust transcriptional competency of the human *FTO* promoter. **c,** In contrast, the 2,820-bp proximal human *IRX3* promoter is not sufficient to drive LacZ expression, which is consistent with an enhancer-dependent transcriptional control mechanism.

a*IRX3* expression in human tissues**b***IRX3* expression in human brain

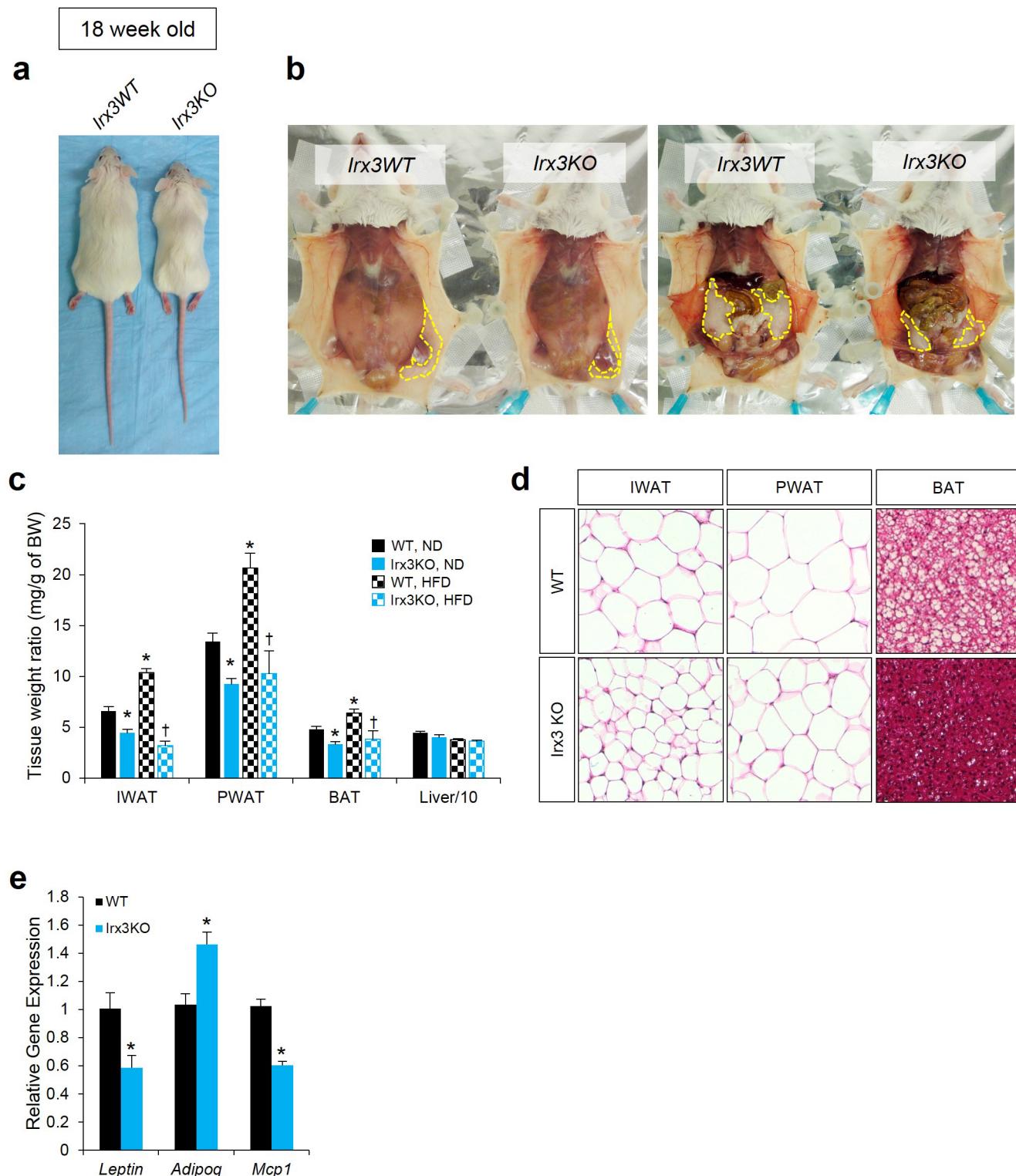
Extended Data Figure 4 | *IRX3* expression in human brain. **a,** *IRX3* expression in human tissues including brain. Expression data, measured on Affymetrix HG-U133 arrays, were obtained from the Body Atlas, Tissues (<http://www.nextbio.com>). The median expression across all 128 human tissues from 1,068 arrays is shown by the red line. **b,** *IRX3* expression in 11

different regions of human brain. Data were retrieved from Human Brain Transcriptome data (<http://www.molecularbrain.org>). Amyg: amygdala; Caud nuc: caudate nucleus; Cere: cerebellum; Corp Call: corpus callosum; DRG: dorsal root ganglion; Frnt Cort: frontal cortex; Hippo: hippocampus; Hypo: hypothalamus; Pit: pituitary; Spine: spinal cord; Thal: thalamus.



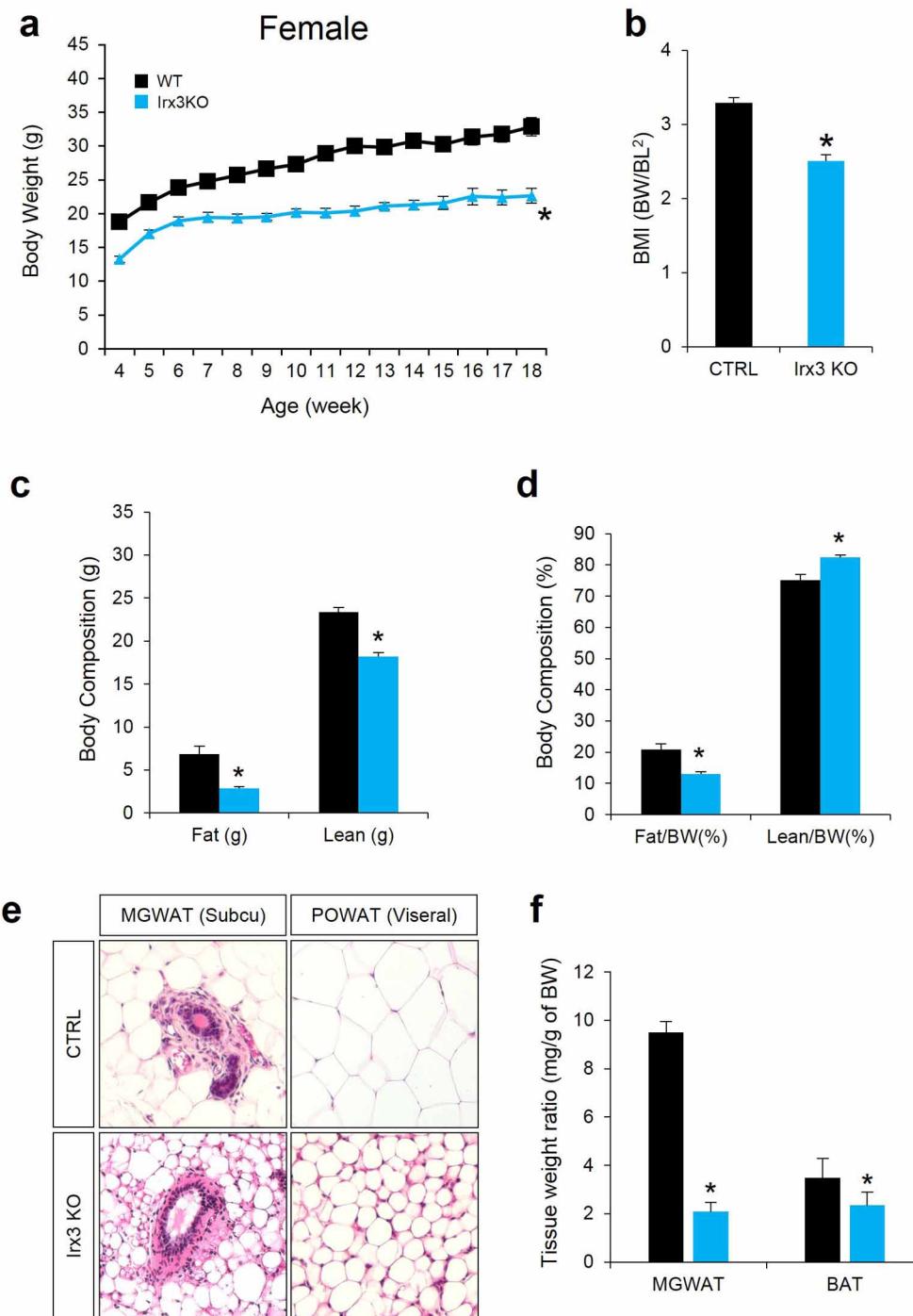
Extended Data Figure 5 | Linkage disequilibrium in *FTO*. Linkage disequilibrium (LD) plot of (logarithm (base 10) of odds) (LOD) score from HapMap phase II European data set, visualized in the UCSC browser. LD blocks are outlined in black. Obesity-associated SNPs from the National

Human Genome Research Institute (NHGRI) GWAS catalogue are shown above, in green, demonstrating why this LD block is considered to define the ‘association interval’.



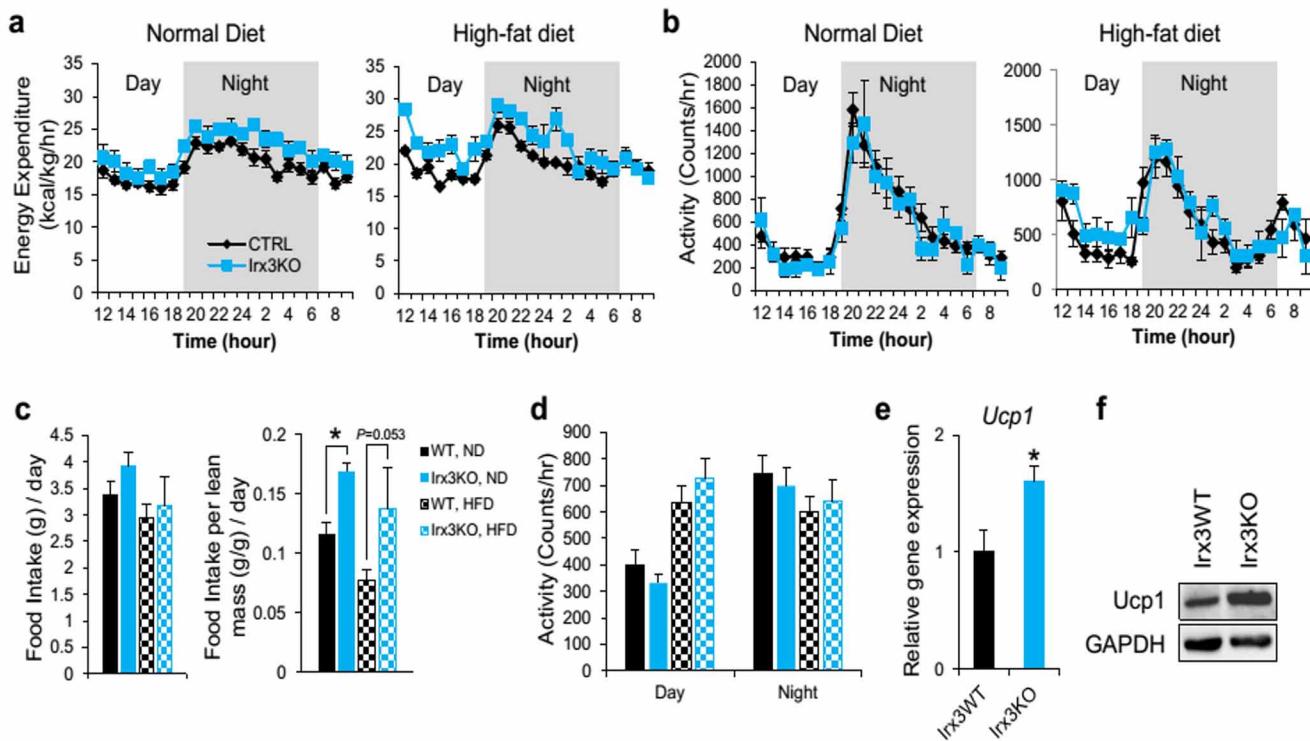
Extended Data Figure 6 | *Irx3*-knockout male mice are leaner with reduced adiposity. **a**, Representative photograph of WT and *Irx3* KO mice fed ND at 18 weeks of age. **b**, Representative anatomical views of WT and *Irx3* KO mice fed ND. Yellow dotted lines depict subcutaneous IWAT (left) and visceral PWAT (right). **c**, Tissue weights as a percentage of body weight showed smaller fat pad sizes in *Irx3* KO mice, compared to WT mice, in both ND and HFD conditions. (ND, WT/KO, $n = 20/12$; HFD, WT/KO, $n = 8/5$.) Data are mean \pm s.e.m.

(* $P < 0.05$ versus WT, ND; † $P < 0.05$ versus WT, HFD). **d**, Representative H&E sections of PWAT, IWAT and BAT from ND mice demonstrated smaller adipocyte size in *Irx3* KO mice than control mice. **e**, Quantitative PCR of WT and *Irx3* KO PWAT for the indicated marker genes: leptin (*leptin*) and adiponectin (*adipoq*) are adipogenic markers, positively and negatively associated with adiposity, respectively; *Mcp1* correlates positively with adiposity. (* $P < 0.05$ versus WT value.) (WT/KO, $n = 10/7$).



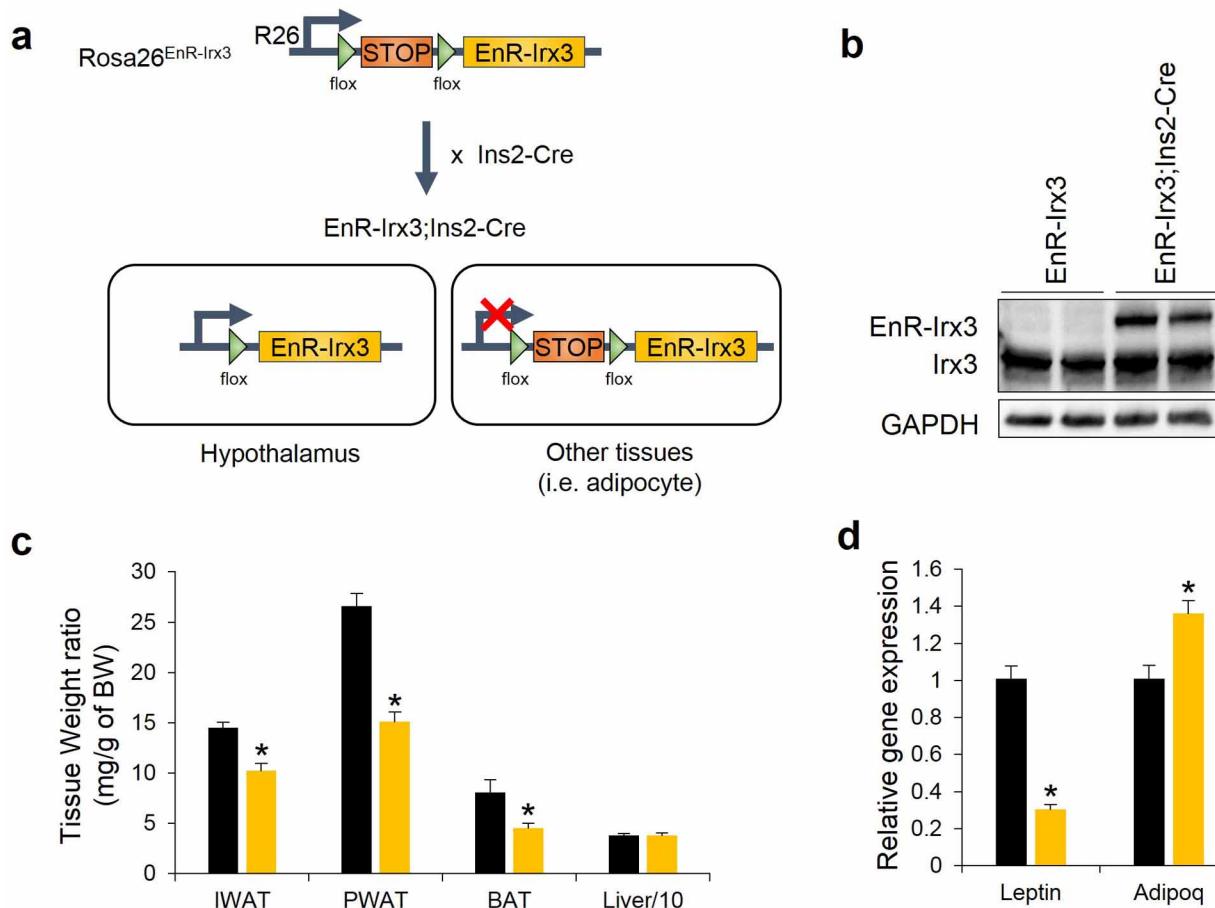
Extended Data Figure 7 | *Irx3*-knockout female mice are leaner with reduced adiposity. **a**, Body weight (BW) changes of WT and *Irx3* KO female mice fed a normal diet (ND). (WT/KO, $n = 15/14$). **b**, BMI, calculated by BW/BL² (BL, body length), is lower in *Irx3* KO female mice. (WT/KO, $n = 7/7$). **c, d**, Body composition analysis showed reduced fat mass and to a lesser extent reduced lean mass in *Irx3* KO female mice compared to WT mice, leading to

decreased fat mass ratio (WT/KO, $n = 9/8$). **e**, Representative H&E-stained sections of mammary gland (MG) WAT and periovarian (PO) WAT revealed smaller adipocyte size in *Irx3* KO female mice, compared to WT. **f**, MGWAT and BAT weights as a percentage of body weight (WT/KO, $n = 4/5$). Data are mean \pm s.e.m. (* $P < 0.05$ versus WT value.)



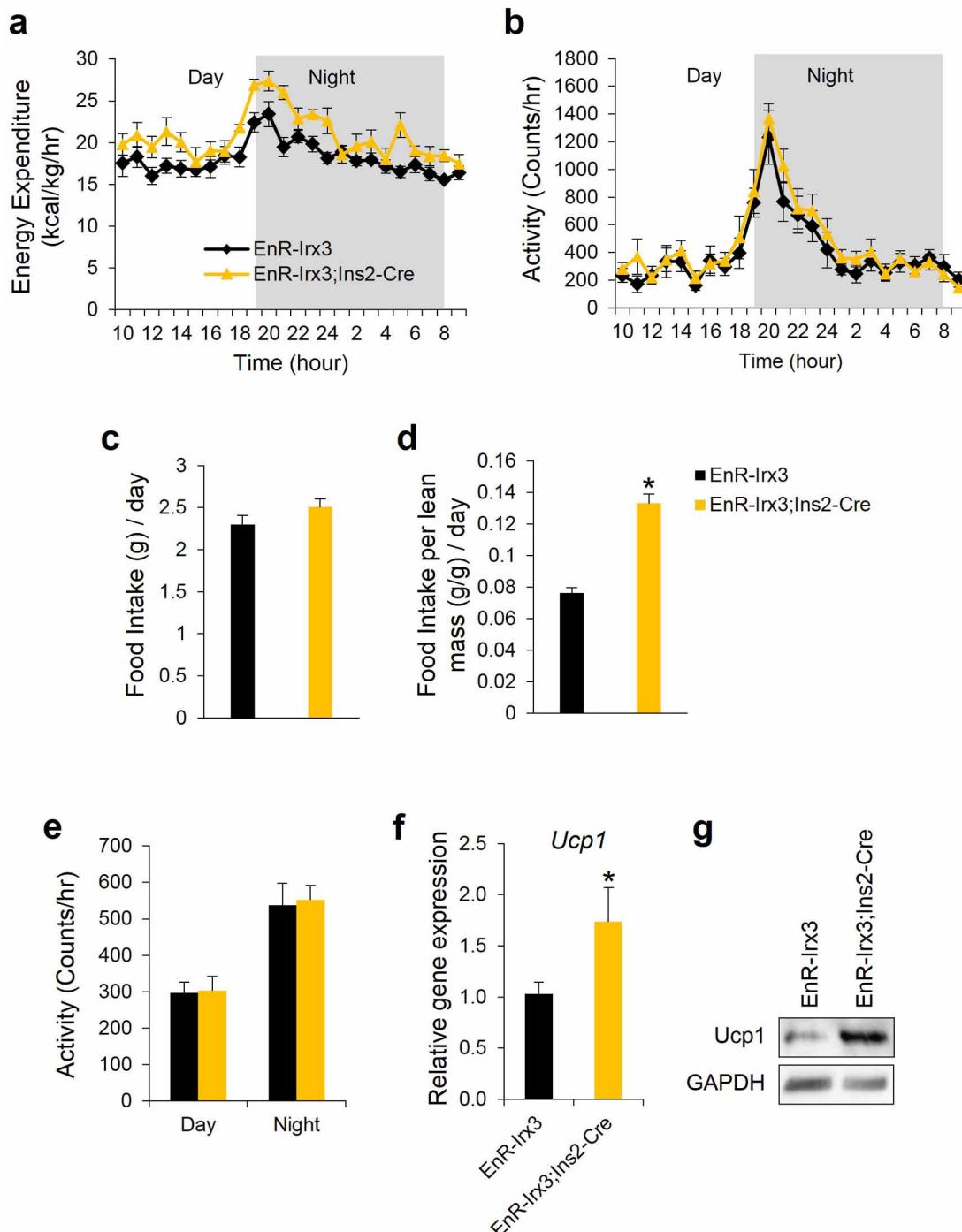
Extended Data Figure 8 | Higher energy expenditure of *Irx3*-knockout mice. **a**, Energy expenditure over a 24-h period, corrected for lean mass ($\text{kcal kg}^{-1} \text{h}^{-1}$), for 18-week-old WT and *Irx3* KO mice fed with ND and HFD (ND WT/KO, $n = 7/5$; HFD WT/KO, $n = 8/4$). **b**, Locomotor activity of WT and

Irx3 KO mice. **c**, Average amount of food intake over a 24-h period with or without normalization to lean mass. **d**, Average locomotor activity measured over 24 h. **e**, **f**, Elevated Ucp1 gene and protein expression in BAT (WT/KO, $n = 7/6$). Data are mean \pm s.e.m. * $P < 0.05$ versus WT value.



Extended Data Figure 9 | Hypothalamic-specific *Irx3* dominant-negative mice are leaner with reduced adiposity. **a**, Schematic diagram of generation of transgenic mice overexpressing dominant-negative *Irx3* in the hypothalamus. **b**, Immunoblot analysis showed *EnR-Irx3* expression in the hypothalamus of mutant mice without affecting endogenous *Irx3* expression, compared to

control mice. **c**, Tissue weights as a percentage of body weight showed that fat pad sizes are smaller in mutant mice, compared to control mice. **d**, Reduced leptin expression and increased adiponectin gene expression in PWAT of mutant mice (control/mutant, $n = 5/7$). Data are expressed as mean \pm s.e.m. * $P < 0.05$ compared to control group.



Extended Data Figure 10 | Higher energy expenditure of hypothalamic dominant-negative *Irx3* mice. **a**, Energy expenditure over a 24-h period, corrected for lean mass ($\text{kcal kg}^{-1} \text{h}^{-1}$), for 18-week-old mice. **b**, Locomotor activity for mice in panel **a**. **c, d**, Average amount of food intake over a 24-h

period with or without normalization to lean mass. **e**, Average locomotor activity measured over 24 h. **f, g**, Elevated gene and protein expression of *Ucp1* in BAT of mutant mice (control/mutant, $n = 5/7$). Data are expressed as mean \pm s.e.m. * $P < 0.05$ compared to control group.