Genome-wide analysis of promoter contacts identifies novel regulators of late-stage adipogenesis

Ionel Sandovici[1,2,3,*,£], Borbala Mifsud[4,5,*,£], Amy Emery[6,$], Pawan Gulati[1,$], Katherine A. Kentistou[7,$], Ayesha Banu[8,$], Niamh Campbell[1,2], Bryn S. Hardwick[6], Alex T. Crooks[6], Denise S. Fernandez-Twinn[1], Lais V. Mennitti[1], Luma Srour[4], Sherine Awad[1], Davide Chiarugi[1], Russell S. Hamilton[3,9], Steven W. Wingett[10,11], Peter Fraser[10,12], Ken K. Ong[7], Stefan Shoenfelder[10,13], Farhan Mohammad[8,%], Stephen O'Rahilly[1,%], John R.B. Perry[1,7,%], Ashok R. Venkitaraman[6,14,%], Susan E. Ozanne[1,%], Miguel Constância[1,2,3,%,£,§]

[1]Metabolic Research Laboratory, Wellcome-MRC Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge CB2 0QQ, UK

[2]Department of Obstetrics and Gynaecology and National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge, CB2 0SW, UK

[3]Centre for Trophoblast Research, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3EG, UK

[4]Division of Genomics and Translational Biomedicine, College of Health and Life Sciences, Hamad Bin Khalifa University, Education City, Doha, Qatar

[5]William Harvey Research Institute, Queen Mary University of London, London, UK

[6]Medical Research Council Cancer Unit, University of Cambridge, Hills Road, Cambridge, CB2 0XZ, UK

[7]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge, CB2 0QQ, UK

[8]Division of Biological & Biomedical Sciences (BBS), College of Health & Life Sciences (CHLS), Hamad Bin Khalifa University (HBKU), Doha, Qatar

[9]Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH, UK

[10]Nuclear Dynamics Programme, The Babraham Institute, Cambridge, CB22 3AT, UK

[11]MRC Laboratory of Molecular Biology, Cambridge, UK

[12]Department of Biological Science, Florida State University, Tallahassee, FL, USA

[13]Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, UK

[14]Cancer Science Institute of Singapore, National University of Singapore, 14 Medical Drive, Singapore 117599, Singapore

*These authors contributed equally

[$]These authors contributed equally

[%]These authors contributed equally

[£]Corresponding authors: Ionel Sandovici (is299@cam.ac.uk), Borbala Mifsud (nagybori@gmail.com) and Miguel Constância (jmasmc2@cam.ac.uk)

[§]Lead author: Miguel Constância (jmasmc2@cam.ac.uk)

SUMMARY

Adipogenesis is a multi-step process, with epigenetic mechanisms and dynamic 3D chromatin folding thought to play important regulatory roles. However, the kinetics and functional roles of promoter contacts during late-stage adipogenesis are unknown. Here, using multi-omics approaches, we found evidence for promoter switching and widespread 3D rewiring of promoter contacts, as well as changes in the transcriptome and epigenome in late-stage adipogenesis. We identified several clusters of promoter contacts with unique temporal profiles suggesting crucial roles for distal enhancers. By integrating transcriptomics, promoter-capture Hi-C and a siRNA screen of druggable genes, we identified 19 novel regulators of late-stage adipogenesis, over half of which have peptidase or ubiquitin-protein ligase activities. Population-based genetic analyses showed that three of the 19 genes (*LAP3*, *CELA1* and *GPR157*) are involved in regulation of adiposity in humans. These findings shed new light on the epigenetic regulation of late-stage adipogenesis, advancing our understanding of the mechanisms that underpin the formation of functional adipocytes and identifying potential targets for preventing/treating obesity and related disorders.

INTRODUCTION

The increasing worldwide prevalence of obesity in recent decades has sparked a growing interest in understanding the mechanisms that control adipocyte differentiation and maturation. Adipogenesis is a multistep process, with an initial phase in which mesenchymal precursors commit to pre-adipocytes, followed by a second phase of differentiation, which involves cell cycle arrest and sustained lipogenesis, to form functional, insulin-responsive mature adipocytes (Ghaben and

Scherer, 2019). Much of our knowledge about the molecular mechanisms driving adipogenesis comes from several cell culture models, such as the murine 3T3-L1 (Green and Kehinde, 1974), and OP9-K cell-lines (Wollins et al., 2006), which differentiate into adipocytes in response to adipogenic cocktails. In the case of the 3T3-L1 cell-line, differentiation takes place in three stages: early (first 24–36 hours), intermediate (hours 36–72) and late (days 3–7) (Ntambi and Young-Cheul, 2000). Mouse lineage tracing models have confirmed that these *in vitro* models recapitulate key steps of adipocyte differentiation *in vivo* (Berry and Rodeheffer, 2013; Stefkovich et al., 2021). However, while we have a detailed view on the molecular events that regulate the early and intermediate stages of adipocyte differentiation, we know substantially less about putative regulators of late differentiation events (and in particular epigenetic mechanisms driving them), leading to the formation of fully mature adipocytes. Adipocyte cell size, turnover and number are major determinants of fat mass, alterations in which are strongly associated with pathology. For example, a number of studies have shown a strong association between large adipocytes and cardiometabolic disorders, including type 2 diabetes risk, as well as associations with dyslipidemia, insulin resistance and hypertension (Rosen and Spiegelman, 2014; Lotta et al., 2017).

Epigenetic regulation plays a key role in cellular differentiation. Many epigenetic changes that modulate transcription during cell differentiation occur at enhancers, which are DNA sequences that activate transcription independent of their location, distance or orientation with respect to the promoters of genes they regulate (Ong and Corces, 2011). H3K4me1 (histone H3 monomethylated at lysine 4) is a key chromatin modification at many enhancers, irrespective of their activity, while H3K27Ac (histone H3 acetylated at lysine 27) and H3K27me3 (histone H3 trimethylated at lysine 27) specify whether an enhancer is active or poised for activation, respectively (Ong and Corces, 2011). The impact of enhancers on promoters is achieved through 3D chromatin folding that allows direct physical contact between them (Schoenfelder and Fraser, 2019). A recent methodological advance based on the classical 3C (chromosome conformation capture) assay (Dekker et al., 2002) – the so-called promoter-capture Hi-C (PCHi-C) enabled mapping of all promoter contacts in a single experiment for both the mouse (Schoenfelder et al., 2015) and the human (Mifsud et al., 2015) genomes. The application of this technique to 3T3-L1 cells demonstrated reorganization of promoter-anchored chromatin loops during early and intermediate stages of adipogenesis, with evidence for substantial rewiring as early as four hours following the addition of the differentiation cocktail (Siersbæk et al., 2017). However, this methodology has not been applied to late-stage adipocyte differentiation.

To address this knowledge gap, in the current study, using novel as well as publicly available data, we surveyed the dynamics of promoter-anchored chromatin loops by PCHi-C, TSS (transcription start

site) activities by CAGE (cap analysis of gene expression)-seq, histone marks by ChIP (chromatin immunoprecipitation)-seq and DNA methylation by whole-genome bisulfite sequencing (WGBS) during 3T3-L1 adipocyte differentiation. Additionally, we performed an *in vitro* siRNA screen of >2,900 druggable genes in the OP9-K cell-line, measuring changes in lipid accumulation upon induction of adipocyte differentiation. Through intersectional analyses of these data sets, we predicted novel regulators of late-stage adipogenesis, and validated selected candidate genes by performing *in vivo* siRNA knockdowns in *Drosophila* fat body and by mining mouse phenotyping (MGI – Mouse Genome Informatics, and IMPC – International Mouse Phenotyping Consortium) and human GWAS (genome-wide association studies) data for associations with measurements of obesity. Our findings expand the list of regulators of adipocyte differentiation with previously unrecognized molecules that influence late-stage adipogenesis.

RESULTS

In vitro adipogenesis is accompanied by extensive transcriptional changes and rewiring of promoter-anchored chromatin loops

We differentiated 3T3-L1 pre-adipocytes (mouse, female sex, see Methods) (Green and Meuth, 1974) for seven days in adipogenic media, which led to uniform accumulation of lipid droplets and transcriptional changes of known marker genes, such as up-regulation of *Adipoq*, *Cebpa* and *Pparg* (Mikkelsen et al., 2010) and down-regulation of *Zfp521* (Kang et al., 2012) (Figure S1A). We measured TSS activities in undifferentiated (day zero – D0) and terminally differentiated (day seven – D7) cells using nAnT-iCAGE sequencing (henceforth referred to as CAGE-seq), as described previously (Murata et al., 2014). We identified over 118,000 peaks (see Methods), corresponding to TSSs, of which one fifth were differentially expressed between D0 and D7 cells (Figure 1A). These differentially expressed peaks correspond to ~7,800 differentially expressed genes (DEGs – Table S1), two-thirds of which were down-regulated during adipocyte differentiation (Figure 1A). To validate this dataset, we performed qRT-PCR for twelve DEGs in independent replicates and found a strong positive correlation between the two methods (Figure S1B). Gene ontology (GO) analysis of down-regulated DEGs showed enrichment of biological processes related to cell cycle, cell division, transcription, DNA repair and cell migration (Figure 1B). Genes up-regulated during adipogenesis were related to biological processes such as lipid metabolism, mitochondrial function and translation (Figure 1B). CAGE-seq analysis also identified over 600 genes that underwent a promoter switch during 3T3-L1 adipogenesis, *i.e.,* genes with at least two transcripts changing expression (differentially expressed peaks) in the opposite direction (Figure 1C; Table S1). This set of genes was

enriched in biological processes related to regulation of transcription, translation and chromatin organization (Figure S1C). Examples of genes that exhibited a promoter switch during adipogenesis include *Akt2*, *Net1* (Figure 1D), *Fhl1*, *Jmjd1c* and *Smarcd2* (Figure S1D).

Next, using FACS, we isolated GFP+ committed pre-adipocytes from the gonadal fat of male *Zfp423*^{GFP} reporter mice (Gupta et al., 2012), as well as matched mature adipocytes and subjected these cells to RNA-seq analyses. We identified 5,174 DEGs *in vivo*, of which ~2,500 were common with DEGs identified by CAGE-seq in 3T3-L1 cells when the same threshold for fold-change was used (Table S1). Over 85% of the common DEGs had the same direction of change during adipogenesis, with a strong positive correlation (Figure S1E). This analysis underlines that the *in vitro* 3T3-L1 model recapitulates well the molecular events taking place during adipogenesis *in vivo*.

To obtain a 3D view on the regulatory regions that interact with all promoters and could lead to changes in gene expression, we then performed PCHi-C in D0 and D7 3T3-L1 cells, as described (Schoenfelder et al., 2015). Using GOTHiC (Schoenfelder et al., 2018), we identified 14,723 differential interactions (DInt) at the resolution of single *Hind*III fragments (see Methods) (Table S2). Only a small proportion of DInt engaged two annotated promoters (*i.e.* promoter-promoter interactions), with >93% joining promoters with non-promoter regions (*i.e.* promoter-other, Figure 1E). These DInt were associated with important changes in the pattern of specific epigenetic marks, both at the promoter and the distal interacting fragments. At promoters, the most dynamic epigenetic marks between D0 and D7 were the repressive histone marks H3K27me3 and H3K9me3, with more subtle changes in levels of H3K4me1 (Figure 1F). The strongest changes observed at the distal interacting fragments (other) were those related to enhancer activity (H3K27ac and H3K4me1) and transcriptional activity (H3K4me3) (Figure 1F). We designed quantitative 3C (q3C) assays (Dekker, 2006) and performed validation in independent D0 and D7 replicates for six predicted DInt, spanning linear distances between 61 Kb and 1.24 Mb. All six q3C tests demonstrated significant changes in the relative interaction frequencies, in the same direction measured by GOTHiC (Figures 1G and S1F). Among these, the interaction between the promoter of *Fabp4*, a critical regulator of intermediate adipocyte differentiation (Bahrami-Nejad et al., 2022), and the intronic fragment at the *Fabp12* locus during adipogenesis has been reported previously (Siersbæk et al., 2017; He et al., 2018). We found that this interaction becomes significantly stronger at D7, and that at the same time the intronic *Fabp12* region gains the epigenetic signature of an active enhancer, with enrichment of H3K27Ac and a putative, actively transcribed eRNA (Figure 1G). Together, these results show that D0 and D7 3T3-L1 cells differ dramatically in their transcriptional output, distribution of epigenetic marks and spatial organization of the genome.

Cluster analysis uncovers sets of promoters with enhancer contacts during late-stage adipogenesis

Having established that D0 and D7 3T3-L1 cells exhibit thousands of DInt, we next used GOTHiC to assess the dynamics of promoter-anchored chromatin loops across distinct stages of adipocyte differentiation (D0, 4 hours, D2 and D7). In this aim, we integrated our PCHi-C data with that generated by Siersbæk et al. (Siersbæk et al., 2017) who analysed earlier stages of adipocyte differentiation, up to D2. K-means clustering identified 23,606 DInt in nine groups with unique temporal profiles, each one including hundreds to several thousands of promoters (Figure 2A; Table S2). Among these, clusters 1, 2, 6 and 8 contain DInt that are transitory during early or intermediate stages of adipocyte differentiation (Figure 2A; Table S2). Clusters 4 and 9 contain DInt that become significantly stronger during terminal adipogenesis, while clusters 3, 5 and 7 include DInt that become significantly weaker in later stages of adipocyte differentiation (Figure 2A; Table S2).

To assess the changes of epigenetic marks at rewiring promoter-anchored chromatin loops during 3T3-L1 adipocyte differentiation, we calculated the enrichment of four histone modifications (H3K27ac, H3K27me3, H3K4me1 and H3K4me3) across the promoters and distal interacting fragments of all nine clusters, based on previously reported ChIP-seq data at D0, D2 and D7 (Mikkelsen et al., 2010). Additionally, we integrated DNA methylation (Park et al., 2022) and H3K9me3 (Matsumura et al., 2015) measurements performed in 3T3-L1 pre-adipocytes and adipocytes (Figures 2B and S2A). The most striking observation from these analyses was the gain in H3K4me1 and H3K27ac, with concomitant loss of H3K27me3 at the distal fragments of cluster 4, particularly between D2 and D7, suggesting enhancer activation during late stages of adipocyte differentiation (Figure 2B). Distal fragments of cluster 9 DInt were associated with a similar, albeit milder pattern of changes, while distal fragments in cluster 7 gained higher levels of H3K9me3 during adipogenesis (Figure 2B). In contrast to the dynamic changes observed at the distal fragments of several DInt clusters, histone marks remained relatively stable at the promoter regions (Figure S2A). DNA methylation levels were different between clusters, at both distal fragments and promoters; however, globally they remained stable during adipocyte differentiation (Figures 2B and S2A).

We then analysed the dynamics of CAGE-seq peaks, per DInt cluster, at distal interacting fragments and promoters. The distal fragments of cluster 4 DInt were the strongest enriched in differential CAGE-seq peaks up-regulated during 3T3-L1 adipogenesis, with smaller, but significant enrichments also observed at clusters 3, 6, 8 and 9 (Figure 2C). Promoters of clusters 4 and 9 were enriched in CAGE-seq peaks up-regulated during adipogenesis, while those of clusters 1, 3, 5, 6 and 7 were

enriched in CAGE-seq peaks down-regulated during adipogenesis (Figure 2D). Next, we analysed gene expression patterns *in vivo*, per DInt cluster and observed a significant up-regulation at mRNA levels in adipocytes versus pre-adipocytes for clusters 4 and 8 (Figure 2E). Notably, analysis of previously published data on gene expression changes during 3T3-L1 *in vitro* adipogenesis (Mikkelsen et al., 2010) showed that genes belonging to clusters 4, 8 and 9 had significant up-regulation of mRNA levels in late stages of adipogenesis (*i.e.* between D2 and D7) (Figure S2B and Table S2).

Using i-cisTarget (Imrichova et al., 2015), we found notable differences in the enrichment of transcription factor (TF) binding motifs at promoters and distal fragments of the nine DInt clusters (Figure S2C and Table S2). Overall, there were fewer TF binding motifs enriched at promoters than at distal fragments for all nine DInt clusters and many TFs enriched at distal fragments were exclusive for one of the nine DInt clusters (Table S2). Among the TF binding motifs enriched exclusively at the distal fragments of only one of the nine clusters associating up-regulated CAGE-seq peaks were CEBPA, CEBPB, CEBPD and CEBPE (cluster 1), GLI1, GLI2, GLI3 (cluster 3), KLF6, KLF13 (cluster 4), PLAGL1 (cluster 5), PML, SIN3A and ZBTB33 (cluster 9) (Figure S2C and Table S2). In contrast, many TFs enriched at promoter regions were shared between several clusters, such as NRF1 (promoters associated with up-regulated CAGE-seq peaks of clusters 3, 4, 6, 8 and 9), E2F4 (clusters 1, 3, 6, 8 and 9) and ELF1 (clusters 5, 6, 7 and 9) (Figure S2C).

Pathway analyses identified clusters 4 and 9 as enriched in genes related to fat differentiation and lipid metabolic processes (Figures 2F and Table S2). Clusters 5, 6 and 7 had enrichment of genes related to response to lipid, while cluster 8 demonstrated an enrichment in genes related to lipid transport and localization (Figure 2F and Table S2). Clusters 1 and 3 had enrichment of genes related to regulation of gene expression and developmental processes, while cluster 2 was void of any significant GO term (Table S2).

Overall, these findings demonstrate that the promoter-anchored chromatin loops are dynamic throughput all stages of adipocyte differentiation. In the case of cluster 4, the chromatin interactions become stronger particularly between D2 and D7, enabling physical proximity between promoters and distal fragments that exhibit epigenetic signatures of active enhancers, concomitant with transcriptional up-regulation during late-stage adipogenesis.

siRNA knockdowns of genes engaged in DInt identify novel regulators of late-stage adipogenesis

To validate the functional role of DInt genes in adipogenesis at larger scale, we next performed a siRNA screen, using The Mouse siGENOME Druggable Genome Library, which contains 2,905 genes considered potential targets for therapeutics (Table S3). We carried out this screen in OP9-K mouse cells (female sex, see Methods), a robust model to rapidly study adipocyte differentiation *in vitro* (Lane et al., 2014). We used 30% as threshold for significant changes in lipid droplet formation at the end of the differentiation time course (see Methods and Figure S3A). Based on this criterion, the screen identified 981 genes that reduced and 41 genes that increased lipid accumulation, respectively (Figure 3A). These sets of genes include several known regulators of adipocyte differentiation, such as *Pparg* and *Pdgfrb* (Ghaben and Scherer, 2019) (Table S3).

Of the total of 2,905 "druggable" genes, 1,216 (~42%) were part of the nine DInt clusters, with 432 showing a significant impact on lipid accumulation upon siRNA knockdown, using the scoring system presented above (siRNA knockdown of 410 genes reduced and 22 genes increased lipid accumulation, respectively; Table S3). However, only cluster 4 demonstrated a significant enrichment in genes that enhanced lipid accumulation upon siRNA knockdown. On the contrary, none of the clusters were enriched in genes that reduced lipid accumulation following siRNA knockdown (Figure 3B). To refine a list of novel putative regulators of late-stage adipogenesis, we used the following criteria: 1) promoters whose interaction frequencies show the most pronounced changes between D2 and D7 (clusters 3, 4, 5, 7 and 9) – which included a total of n=13,778 DInt, engaging 7,346 promoters; 2) significant expression changes by CAGE-seq (>2 fold) between 3T3-L1 pre-adipocytes and adipocytes (n=7,807 genes); 3) differential expression (>1.5 fold) by RNA-seq between primary mouse pre-adipocytes and adipocytes (n=7,596 genes); 4) genes whose depletion induced >30% changes in lipid droplet formation in OP9-K cells (n=1,022 genes). This intersectional search led to the identification of 49 gene promoters engaged in 122 DInt (Figure 3C and Table S3). We manually curated this list (see Methods) and identified 19 genes without any previously known link to adipogenesis (Figure S3B), over half of which have peptidase (*Cela1*, *Lap3*, *Prss23*, *Clpx*) or ubiquitin protein ligase (*Znrf2*, *Rnf125*, *Rnf139*, *Hectd2*, *Fbxo17*, *Fbxl14*, *Trim21*, *Fbxw5*) activities (Table S3).

Drosophila melanogaster has been used successfully in the past to screen for genetic regulators of adipocyte differentiation (Pospisilik et al., 2010). Within the lists of 49 and 19 genes identified through the intersectional analyses presented above, we found 13 and 5 that have strong homology between mouse and *Drosophila* (FlyBase score >12/15; Table S3), respectively. We used the pumpless (*ppl*)-GAL4 driver (Colombani et al., 2003) to perform *in vivo* siRNA-mediated knockdown specifically in the fat body of *Drosophila* starting in the larva stage and throughout development (Figures S3C and S3D). The fat body is the organ that has storage and humoral functions associated

with nutrition, comparable to that of vertebrate adipose tissue and liver (Colombani et al., 2003). We selected for this *in vivo* analysis three genes (Figures 3D-H and S3D-G): two novel candidates (*Clpx* and *Fbxl14*), and one with a previously known role in adipogenesis (*Psma1*) (Table S3) and performed lipid accumulation measurements in adult flies. Knockdown of two of the three genes (*Psma1*/Prosalpha6T in *Drosophila* and *Clpx*/ClpX) led to significant changes in lipid accumulation in flies' fat bodies, in the same direction with the observations made in OP9-K cells (Figures 3D–H), while the third gene (*Fbxl14*/Ppa) led to opposite effects (Figures S3E–G).

Thus, by using two siRNA knockdown systems, one *in vitro* (OP9-K cells) and one *in vivo* (*Drosophila* fat body) we identified and validated several novel regulators of late-stage adipogenesis predicted by our omics analyses.

Regulators of late-stage adipogenesis harbour significant associations with measurements of obesity in mouse and human

To gather additional evidence for links between regulators of late-stage adipogenesis and measurements of obesity, we performed a search in two publicly available databases that contain phenotypes for thousands of mouse mutants (MGI – phenotypes available for 13,487 mutant genes and IMPC – 7,577 mutant genes phenotyped), as well as in the central human GWAS catalogue (see Methods). First, we identified three mouse phenotypes of interest: increased fat amount, decreased fat amount, and abnormal fat morphology (see Methods). Within the MGI dataset, which contains both loss-of-function and gain-of-function mutations, we observed greatest enrichment in effects across all three phenotypes among genes in cluster 4 (Figure 4A and Table S3). Other clusters exhibited less pronounced effects: increased fat amount (cluster 8), decreased fat amount (clusters 3, 5 and 9) and abnormal fat morphology (cluster 5) (Figure 4A and Table S3). Within the IMPC dataset, which contains only loss-of-function mutations, we observed higher risk for decreased total body fat amount at cluster 5 and higher odds ratios for abnormal adipose tissue morphology at clusters 5 and 9 (Figure S4A and Table S3). Notably, 8/49 and 2/49 genes identified through the intersectional analysis showed an association with fat-related phenotypes in the MGI and IMPC databases, respectively (Table S3). We also sought to determine whether there was any relationship between the nine DInt clusters and various measures of human adiposity identified through GWAS. This analysis identified DInt clusters with significant enrichment of genes associated to two specific adiposity-related GWAS traits, namely "adipose tissue measurement" and "obesity" in clusters 3, 5, 8 and 9 (Figure 4B). Additionally, we found significant enrichment of genes for the two specific

GWAS terms, analysed separately (clusters 1, 4, 5, 8 and 9 for adipose tissue measurement; clusters 3, 7 and 8 for obesity) (Figure S4B).

Comparative studies between species enable identification of key regulatory regions that were conserved during evolution (Peters et al., 2007). Therefore, our next analysis searched for DInt belonging to clusters 3, 4, 5, 7 and 9 (*i.e.* strong changes of interaction frequencies in late stages of adipogenesis), in which the promoter and the distal interacting fragment maintained the synteny between mouse and human. We then selected those DInt that contain human GWAS obesity-related SNPs (single nucleotide polymorphisms) at the human region homologous to the distal mouse fragment. We identified four such cases. In the case of rs912056, located on human chromosome 6 (reported gene is *LY86*), we observed a cluster 5 DInt between the homologous mouse fragment and the promoter of *Rreb1*, a gene that showed significant mRNA up-regulation during adipocyte differentiation (Figures 4C and 4D). *Rreb1* was not part of our siRNA screen performed in the OP9-K cells. However, *Rreb1* has a strongly conserved homologue in *Drosophila* (namely peb, FlyBase score of 12/15). Using the *ppl*-GAL4 driver described above, we performed siRNA knockdown of peb, and found a significant reduction of lipid accumulation in the fat body (Figure 4E). Additionally, IMPC data showed a significant reduction of fat content in males heterozygous for a mutant *Rreb1* allele (Figure S4C). Based on the experimental evidence obtained, our data suggests that *RREB1* is likely a better candidate to explain the association with the fat distribution traits (ratio of visceral to subcutaneous adipose tissue volume, and ratio of visceral to subcutaneous adipose tissue volume adjusted for BMI – body mass index) reported for rs912056 (Chu et al., 2017). A second example suggesting a novel candidate gene for obesity is that of rs217669, located on human chromosome 14 (reported gene is *SYT16*). The homologous mouse region located on chromosome 12 is engaged in a cluster 7 DInt with the promoter of a nearby non-coding RNA (*1700086L19Rik*, homologous with the human non-coding gene *LINC00643*) (Figure S4D). In two other cases, the candidate genes identified using our synteny approach coincided with the previously reported causal genes. These are *COL14A1* located on human chromosome 8 (SNPs is rs10955960; the homologous mouse fragment on chromosome 15 is engaged in a cluster 5 DInt) and *MAP2K7* located on chromosome 19 (SNPs is rs4804833; homologous mouse fragment on chromosome 8 is engaged in a cluster 9 DInt), thus strengthening the previous GWAS observations (Figure S4D).

To evaluate further a potential link between naturally occurring alleles influencing the function of the 19 novel regulators of late-stage adipogenesis and measurements of obesity, we next interrogated phenotypic and genetic data from recently available large-scale population studies (see Methods). We focused on two measurements of obesity and body fat distribution, namely BMI and WHRadjBMI (waist-hip ratio adjusted for BMI) in sample sizes up to up to 806,834 individuals (see

Methods). We found that three of the 19 genes were located at less than 500kb distance from a genome-wide significant signal for one of these two traits (Table S3): *LAP3* and *CELA1* for BMI (Figure 4F) and *GPR157* for WHRadjBMI (Figure 4G). To link more directly these proximal associated genetic variants to the function of these candidate genes, we then undertook variant-to-gene mapping approaches (see Methods), including assessment of expression quantitative trait loci (eQTL) data and integration of activity-by-contact enhancer maps. These analyses showed that variants that associate with changes in BMI at the *LAP3* and *CELA1* loci might do so via changes in the genes' transcript levels and that the signal at the *GPR157* locus resides within an enhancer of this gene (Table S3).

Collectively these findings demonstrate that our genetic analyses guided by DInt were able to confirm or reassign genes linked to GWAS SNPs associated with obesity and to identify novel genes that exhibit significant associations with measurements of obesity in the mouse or influence population-based variation in phenotypes relevant to obesity or body fat distribution in human.

DISCUSSION

Obesity poses a heterogeneous risk for cardiometabolic complications. The mechanisms that uncouple adiposity from its cardiometabolic comorbidities are complex and include both genetic and environmental factors (James et al., 2021). A deficit or an excess of fat cells (especially hypertrophic fat cells) are both detrimental for health (Rosen and Spiegelman, 2014). Several recent studies provided evidence for the existence of genetic and epigenetic determinants of limited adipose storage capacity (Lotta et al., 2017; Kim et al., 2023) and regional fat distribution (Huang et al., 2021; Agrawal et al., 2022; Akbari et al., 2022; Agrawal et al., 2023) that modulate the risk for insulin resistance. Collectively, these studies are consistent with a recently put forward hypothesis stipulating that subtle forms of lipodystrophy contribute to the risk for cardiometabolic disease at a population level (Mann and Savage, 2019). They also suggest the existence of molecular mechanisms that control late-stage adipogenesis, which, when dysfunctional, lead to the development of adipocytes with reduced capacity to store fat, thus increasing the risk of metabolic disease.

Evidence that long-range chromatin interactions between gene promoters and distal regulatory elements play key roles in transcriptional regulation during cellular differentiation has emerged (Javierre et al., 2016; Rubin et al., 2017; Choy et al., 2018; Zhang et al., 2020), including during early and intermediate stages of *in vitro* adipocyte differentiation (Siersbæk et al., 2017). Based on the previous observations, we hypothesized that cataloguing the promoter-bound chromatin loops with

significant changes in their interaction frequencies at the transition between immature (D2) and mature (D7) adipocytes, could be an efficient way to guide the discovery of novel regulators of late-stage adipogenesis. Our multi-omics analyses, in combination with a siRNA screen of druggable genes in a cellular model of adipogenesis, *in vivo* validation in a *Drosophila* model, and mining of mouse phenotyping and human GWAS datasets, led to the identification of 49 regulators of late-stage adipogenesis, of which 19 have not been implicated in fat differentiation (Table S3). This set of 19 novel regulators was enriched in genes encoding proteins with ubiquitin protein ligase activity or belonging to the SCF (Skp1 – cullin 1 – F-box) ubiquitin ligase complex. This included *Znrf2*, *Rnf125*, *Rnf139*, *Hectd2*, *Fbxo17*, *Fbxl14*, *Trim21* and *Fbxw5.* Moreover, the list of 19 genes also included other members related to ubiquitination: *Traf2*, an essential constituent of several E3 ubiquitin-protein ligases that regulates activation of NF-κB and JNK (with central roles in regulation of cell survival and apoptosis), and *Zfand6*, encoding a zinc finger protein shown to bind TRAF2 and involved in TRAF2-mediated NF-κB signalling (Chang et al., 2011). Ubiquitin ligases are the final and potentially the most important determinant of substrate specificity in the ubiquitination of proteins. Interestingly, with the exception of *Znrf2*, siRNA knockdown of the remaining seven genes with ubiquitin ligase activity reduced lipid accumulation in OP9-K cells, suggesting an important physiological role of the ubiquitin proteasome pathway in the final stages of adipocyte differentiation. Other previously unrecognized regulators of adipogenesis included several proteases: CLPX, a subunit of a major mitochondrial protease complex, thought to stimulate mitochondrial unfolded protein response in mammalian cells; PRSS23, a secreted member of the trypsin family of serine proteases, with proteolytic functions linked to IGFBPs; CELA1, a serine protease that hydrolyses many proteins in addition to elastin; and LAP3, a cytosol leucine aminopeptidase with putative roles in the control of cellular redox status via metabolism of glutathione and in the degradation of glutathione S-conjugates. Our data provides the first evidence for a role of ubiquitin protein ligases and proteases in the regulation of late-stage adipocyte differentiation capacity. Taken together, we found experimental support for a number of new ”druggable” regulators of adipogenesis related to protein homeostasis (“proteostasis”), specifically in pathways related to the degradation of intracellular proteins (the ”proteasome”) and ubiquitination as a marker for regulated proteolysis.

We validated several of the 19 novel candidates through siRNA in *Drosophila* fat body and mining population-level genetic datasets, which brought additional confirmatory evidence for *Clpx*, *LAP3*, *CELA1* and *GPR157* as potential regulators of late-stage adipogenesis. *GPR157* encodes for an orphan receptor that is predicted to be located in the ciliary membrane, with very little known about its function. Primary cilia have emerged recently as pivotal signalling hubs, with key roles in

adipogenesis (Yamakawa et al., 2021). We also gathered additional evidence for several genes previously liked with adipogenesis. Of these genes, *RREB1* has been associated before with increased visceral fat accumulation in human (Chu et al., 2017). Our findings add new members to the list of genes identified in previous studies as regulators of early (*Fos*, *Jun*, *Myc*, *Cebpb* and *Cebpd*) and intermediate (*Pparg, Cebpa* and *Fabp4*) stages of adipogenesis (reviewed by Ntambi and Young-Cheul, 2000; Ghaben and Scherer, 2019). Future studies aimed at identifying additional factors that control late-stage adipogenesis and to understand the signalling or metabolic pathways that they control will be of great interest.

Our study also provided new knowledge on the epigenetic regulation of adipocyte differentiation. One of the findings highlighted by the CAGE-seq analysis was the evidence for promoter switching of over 600 genes during 3T3-L1 adipogenesis. The usage of alternative promoters has been described before in the context of cell fate specification (Molyneaux et al., 2015; Edupuganti et al., 2017), immediate early response (Vacca et al., 2018), cell-cycle-dependent transcription regulation (Wragg et al., 2020), as well as a mechanism by which ubiquitously expressed genes participate in cell-specific functions (Feng et al., 2016). The mechanisms responsible for the alternative promoter usage during adipogenesis require further study. However, as these genes are enriched in GO terms related to regulation of transcription, translation and chromatin organization, they may have a widespread role in guiding adipogenesis. Another important finding in our study is that some clusters containing DInt with the strongest changes between D2 and D7 of adipocyte differentiation have features of enhancers at the distal interacting fragments. This is particularly evident for clusters 4 and 9 that gain two histone marks of active enhancers (H3K27ac and H3K4me1), have enrichment in up-regulated CAGE-seq peaks, suggestive of eRNAs, and are enriched in genes up-regulated during adipogenesis *in vitro*. Clusters 4 and 9 also demonstrate enrichment of genes related to fat cell differentiation and lipid metabolism. Conversely, distal fragments of cluster 7 become enriched in H3K9me3, a suppressive histone mark associated with formation of facultative heterochromatin during cell differentiation (Nicetto and Zaret, 2019), as well as for the genome compartmentalization (Nichols and Corces, 2021), suggesting a potential role in the weakening of the DInt during late-stage adipogenesis that we observed for this cluster. We also identified enrichment of TF binding sites at promoters and distal fragments engaged in DInt. In the case of the TF binding sites enriched at distal fragments, many exhibited specificity for a particular cluster. Interestingly, distal fragments of cluster 1, comprised of promoter-bound DInt that are transitory during early stages of adipocyte differentiation, are enriched in binding sites for several known regulators or early adipogenesis, such as CEBPB and CEBPD (Ntambi and Young-Cheul, 2000). Distal fragments of cluster 4, containing promoter-bound DInt that become strongest in late stages of adipocyte differentiation, are enriched

in binding sites for two known adipogenesis regulators, KLF6 (Li et al., 2005) and KLF13 (Jiang et al., 2015). Therefore, based on the more substantial changes observed at "other ends" than at promoters, we propose that events that take place at distal fragments of DInt, such as changes in epigenetic marks, transcriptional activity (eRNAs) and binding of TFs play a major role in guiding each stage of adipocyte differentiation.

We acknowledge that our study has several limitations. Studies of human adiposity traits uncovered the existence of depot-specific genetic determinants (Schleinitz et al., 2014; Loh et al., 2015; Agrawal et al., 2022), as well as an important influence of sex and age (Heid et al., 2010; Winkler et al., 2015; Song et al., 2018; Rask-Andersen et al., 2019; Link et al., 2020; Akbari et al., 2022). This complexity was not captured by the two cell lines used in this study. Detailed functional characterization of late-stage adipogenesis regulators will require development of new *in vitro* models capable of capturing the complexity of the observations made in human studies. Finally, we acknowledge that not all DInt identified in 3T3-L1 cells are conserved in human, particularly the more distant ones, due to chromosomal rearrangements that occurred during evolution. The application of the PCHi-C assay to a human model of *in vitro* adipogenesis in future studies is warranted to uncover additional novel regulators of late-stage adipogenesis that may play specific roles in our species.

In conclusion, we revealed an adipogenesis-driven promoter switch for a set of genes related to epigenetic regulation and demonstrated that rewiring of promoter-anchored chromatin loops takes place throughout all stages of adipogenesis, enabling the interaction with distal enhancer regulatory elements. Analysis of promoter contacts that demonstrate strong changes at the transition between immature and mature adipocytes is an efficient way to uncover novel regulators of late-stage adipogenesis, as we demonstrated through genetic screens in both *in vitro* and *in vivo* model systems and through population-based genetic analyses. Our approach to focus on discovering druggable targets raises the possibility that our findings could rapidly pave the way for novel approaches in the prevention and treatment of obesity and its related comorbidities.

METHODS

3T3-L1 adipocyte differentiation

Mouse 3T3-L1 cells were grown at 37°C and 5% CO2 in Dulbecco's Modified Eagle's Medium (DMEM high glucose, Sigma-Aldrich #D6546) supplemented with 10% calf serum (Gibco #16010159), 2% glutamine (Sigma-Aldrich #G7513) and 1% Pen-Strep (Sigma-Aldrich #P0781). Cells were induced to differentiate two days post confluency (defined as day zero – D0) by addition of differentiation

media: DMEM high glucose supplemented with 10% fetal bovine serum (Sigma-Aldrich #F7524), 1% Pen-Strep, 1 µg/ml insulin (Sigma-Aldrich #I9278), 390 ng/ml dexamethasone (Sigma-Aldrich #D4902), 1,115 µg/ml 3-Isobutyl-1-methylxanthine (Sigma-Aldrich #I5879) and 2 µM rosiglitazone (Sigma-Aldrich #R2408). Two days after the induction of differentiation, fresh differentiation media supplemented with 1 µg/ml insulin was added. From day four and onward, cells were maintained in differentiation media.

To quantify efficiency of 3T3-L1 differentiation into adipocytes, staining with Oil red O was performed, as previously described (Ramírez-Zacarías et al., 1992). Briefly, cells were fixed in 10% formaldehyde in phosphate buffer (PBS) for one hour, washed with 60% isopropanol, and stained with Oil Red O solution (Sigma-Aldrich #102419) for 10 minutes followed by four washing steps with water, counterstaining of nuclei with hematoxylin (Sigma-Aldrich #105175) and imaging with an optical microscope. Counting the percentage of cells containing Oil Red O positive lipid droplet was performed by ImageJ software (National Institutes of Health, Bethesda MD, USA). Over 90% of cells were Oil Red O positive after seven days of differentiation, using the above protocol.

An additional method to quantify efficiency of 3T3-L1 differentiation into adipocytes was by measuring mRNA levels of known markers of adipogenesis by qRT-PCR. Total RNA was extracted from D0 and D7 3T3-L1 cells using an RNeasy Plus Mini Kit (Qiagen #74134). RNA concentration was measured by NanoDrop (Thermo Fisher Scientific) and quality was assessed in agarose gels. Reverse transcription was performed using the RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific – K1622). qRT-PCR was performed with the SYBR Green JumpStart Taq Ready Mix (Sigma – S4438) and custom-made primers (Table S4) using an ABI Prism 7900 system (Applied Biosystems). For gene expression normalization, we used two housekeeping genes (*Ppia* and *Rplp0*). Levels of expression were calculated using the $2^{-\Delta\Delta Ct}$ method.

CAGE-seq analysis of D0 and D7 3T3-L1 cells

CAGE-seq analysis was performed using the nAnT-iCAGE method, as described (Murata et al., 2014). Briefly, total RNA was extracted from four D0 and four D7 biological replicates of 3T3-L1 cells, using an RNeasy Plus Mini Kit (Qiagen #74134). RNA concentration and quality were measured by NanoDrop (Thermo Fisher Scientific) and an Agilent RNA 6000 Nano Kit (Agilent #5067-1511), respectively. All samples had RIN>9. Library preparation was performed at the Genome Network Analysis Support Facility, RIKEN CLST, with 50-base single read sequencing performed on an Illumina HiSeq2500 instrument. Reads were mapped using TopHat, and RECLU (Ohmiya et al., 2014) was used

to call differentially expressed top peaks and bottom peaks corresponding to reproducible TSSs. Gene level expression was calculated from the RECLU output using top peaks. First, D0 peak level expression was obtained using the formula (2^logConc*2)/(1+2^logFC), then D7 expression values were calculated by D0*2^logFC. The expression values of top peaks annotated to the same gene were summed up for gene level expression. Genes with more than two differential top peaks, where the direction of change was divergent across the differential peaks, were identified as genes involved in promoter switch. Functional analysis was performed using DAVID (Database for Annotation, Visualization and Integrated Discovery; v6.8 https://david.ncifcrf.gov). Enriched gene ontology (GO) terms with FDR < 5% were considered significant. These terms were then clustered semantically using REViGO (Reduce and Visualize GO) (Supek et al., 2011), which removes redundancy, and ordered according to the $\log_{10}$ p values.

RNA-seq analysis of primary pre-adipocytes and adipocytes

Pre-adipocytes and adipocytes were isolated from epididymal fat pads of young (three months-old) hemizygous Zfp423^{GFP} male mice (Gupta et al., 2012). RNA was extracted using the RNeasy Plus Micro Kits (Qiagen #74034) and used to prepare RNA libraries utilizing the TruSeq Stranded mRNA Kit (Illumina #20020595) and sequenced as single-end 50 bp reads using an Illumina HiSeq 4000 platform.

RNA-seq reads were mapped to GRCm39 version of the mouse reference genome sequence using STAR (v2.5.1b) (Dobin et al., 2013). Reads were considered mapped if the similarity was greater than 95% over at least 90% of the read length, as previously described (Haak et al., 2018). FeatureCounts (v1.5) (Liao et al., 2013) was applied for the generation of count tables based on the mapping files. Raw counts were subjected to differential gene expression analysis via DESeq2 (Love et al., 2014) and normalized to CPM (counts per million).

Promoter-Capture Hi-C (PCHi-C) analysis

PCHi-C was performed in triplicate for D0 and D7 3T3-L1 cells as previously described (Schoenfelder et al., 2015). Briefly, 30-40 million cells/sample were fixed with 2% formaldehyde (Agar Scientific #AGR1026) for 10 min at room temperature, prior to harvesting from the tissue culture flasks. Digestion of cross-linked chromatin with *Hind*III (NEB #R3104), labelling with biotin-14-dATP (Invitrogen #19524016) and DNA Polymerase I (Large Klenow Fragment; NEB #M0210) and ligation (with T4 DNA ligase; Invitrogen #15224025) were all performed on nuclei kept intact. Chromatin

cross-linking was then reversed by incubation with proteinase K (Roche #03115879001) at 65°C overnight and the DNA was purified by phenol/chloroform extraction. Aliquots of the Hi-C libraries were used to verify efficiency of chromatin digestion and digestion using standard PCR and qPCR, with primers listed (Table S4). Then, biotin from un-ligated DNA ends was removed using T4 DNA polymerase (NEB #M0203) and the DNA was sonicated (Covaris E220) to an average size of around 400 base pairs. Sonicated DNA was end-repaired using DNA Polymerase I (Large Klenow Fragment; NEB #M0210), T4 DNA polymerase (NEB #M0203) and T4 DNA polynucleotide kinase (NEB #M0201), then dATP was added to the 3' ends of the DNA using Klenow exo- (NEB #M0212). The DNA was then subjected to double-sided SPRI bead size selection (AMPure XP beads; Beckman Coulter #A63881) and biotin-marked ligation products were isolated using MyOne Streptavidin C1 Dynabeads (Invitrogen #65002). After adapter ligation (Illumina PE PCR 1.0 and PE PCR 2.0 primers), the bead-bound Hi-C DNA was amplified with seven PCR amplification cycles. Promoter capture Hi-C was performed using a custom-made RNA capture bait system (Agilent Technologies) consisting of 39,021 individual biotinylated RNAs targeting the ends of 22,225 promoter-containing mouse *Hind*III restriction fragments, as described (Schoenfelder et al., 2015). After a post-capture PCR (four amplification cycles using Illumina's PE PCR 1.0 and PE PCR 2.0 primers), the PCHi-C libraries were purified with AMPure XP beads (Beckman Coulter #A63881) and paired-end sequenced (HiSeq 1000, Illumina) at the Babraham Institute Sequencing Facility.

Quality control of raw fastq files was performed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimming was done when required using TrimGalore (https:://github.com/FelixKrueger/TrimGalore). Reads (220 – 282 million paired-reads/sample) were mapped to the mm9 genome and filtered using the HiCUP pipeline (v0.5.8), which removes experimental artefacts, such as circularised reads and re-ligations, and duplicated reads (Wingett et al., 2015). After de-duplication, the number of valid unique di-tags varied between 90 – 154 million per sample (*i.e.*, between 92.8% and 99.5% of the valid pairs). GOTHiC (Schoenfelder et al., 2018) was used to identify significant differential interactions (DInt) by comparing our own D0 and D7 data, as well as that obtained previously at D0, 4h and D2 (Siersbæk et al., 2017). DInt were taken forward when they overlapped in at least two replicates and $\log_2$ fold changes were calculated as the average between the replicates in which the interaction was classified as DInt. Interactions overlapping with two regions of known artefacts (chr10:106613366-107858706 and chr10:116174799-118176364) were removed.

q3C assays

3C (chromatin conformation capture) assays were performed in independent biological replicates of D0 and D7 3T3-L1 cells. Briefly, 10-15 million cells per sample (6-9 D0 samples and 7 D7 samples) were fixed with formaldehyde, digested with HindIII and ligated with T4 DNA ligase as described above, without the incorporation of biotin-14-dATP. After reversal of chromatin crosslinking, the purified DNA was used in qPCR reactions using 100 ng DNA as template/reaction, the SYBR Green JumpStart Taq Ready Mix (Sigma – S4438) and custom-made primers (Table S4) using an ABI Prism 7900 system (Applied Biosystems). Interaction frequencies were calculated using the $2^{-\Delta\Delta Ct}$ method and normalized against the interaction between two consecutive *Hind*III fragments at the *Rplp0* locus, used as internal control for efficiency of *Hind*III chromatin digestion and DNA ligation.

Clustering of DInt

Interactions that were significant in at least one of the three comparisons (4h vs D0, D2 vs D0, D7 vs D0) were used for clustering. Log_2 fold changes of the interactions were collected from all three comparisons and grouped using k-means clustering (Hartigan and Wong, 1979). The number of clusters was determined using the elbow method by calculating the within group sum of squares for k=2 to k=15. The optimal number of clusters was 9. The profile of the log_2 fold changes of differential interactions was visualized using ggplot2 showing the mean log_2 fold changes and error bars indicating the standard deviation within the cluster.

ChIP-seq and WGBS data processing and enrichment calculation

ChIP-seq data for H3K27ac, H3K27me3, H3K4me1 and H3K4me3 at D0, D2 and D7 was obtained from Mikkelsen et al. (Mikkelsen et al., 2010). DNA methylation data was used from Park et al. (Park et al., 2022) and H3K9me3 from Matsumura et al. (Matsumura et al., 2015). For H3K27ac, H3K27me3, H3K4me1 and H3K4me3 we made use of the peaks identified by Mikkelsen et al. The H3K9me3 ChIP-seq data quality control was performed using FastQC, and reads were trimmed using TrimGalore (https:://github.com/FelixKrueger/TrimGalore). Mapping to the mm9 genome was achieved using BWA-MEM (arXiv:1303.3997) and broad peaks were called using MACS2 (Zhang et al., 2008). We overlapped histone modification peaks with promoter or distal fragments belonging to the nine clusters or the differential interactions between D0 and D7, and labelled each fragment whether it carried the mark or not. Enrichment of fragments with the mark was calculated across clusters/categories and was normalized for the total number of peaks found for that modification at that time point. For DNA methylation, average methylation level was calculated for each *Hind*III

fragment in the genome and enrichment in a cluster/category was compared to the average across them.

Expression microarray analysis in 3T3-L1 and OP9-K cells

OP9-K is an embryonic stromal cell-line of unspecified sex (www.cellosaurus.org). To establish the sex of origin of these cells, we compared gene expression levels inOP9-K cells (Lane et al., 2014) with those of male white adipose tissue (C3HeB/FeJ), as well as of male and female liver samples, retrieved using the accession numbers GSE197101 and GSE176226, respectively. Microarray data normalization was conducted using affy, oligo, and limma packages in R. Then, data were annotated using the annotation table obtained from GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6246) in order to identify the chromosomal location of each gene. Normalized expression values were averaged for replicates and expression levels in the OP9-K cells were compared to the expression in male and female tissues across all chromosomes and separately on the Y chromosome. These analyses revealed that OP9-K cells originated from a female embryo.

3T3-L1 was reported as a spontaneously immortalized cell-line isolated from a male embryo (www.cellosaurus.org). However, recent karyotyping analyses revealed the presence of two X chromosomes (Oh et al., 2021). To establish the sex of 3T3-L1 cells, microarray expression data retrieved using the accession number GSE20752 was normalized with other male and female data. Male white adipose tissue and female mammary gland expression data were taken from GSE10246. All the raw expression data were normalized using affy and limma packages in R. Normalized expression values were averaged across the replicates. Then, the probes were annotated using the GPL1261 annotation table from GEO (https://ftp.ncbi.nlm.nih.gov/geo/platforms/GPL1nnn/GPL1261/annot/) to identify and compare the expression level of genes located on each chromosome and specifically on chromosome Y between the 3T3-L1 cell lines and the known gender tissues. Our analyses are consistent with recent findings suggesting origin in a female embryo.

To compare mRNA expression levels at D0, D2 and D7 in 3T3-Le cells, data generated using GeneChip arrays (Affymetrix) were retrieved from Table S2 of Mikkelsen et al. (Mikkelsen et al., 2010). Data was sorted per DInt cluster, with a coverage between varying between 59.2% of genes (cluster 7) and 71.8% (cluster 4). Data was analysed per individual DInt cluster using one-way ANOVA tests followed by Tukey's multiple comparisons tests.

siRNA screen for regulators of adipocyte differentiation in the OP9-K cells

Mouse OP9-K cells were grown at 37°C and 5% CO2 in OP9 propagation media: MEM-α (Life Technologies #12571-063), supplemented with 20% FBS (Sigma-Aldrich #F4135), 2mM L-glutamine (Sigma-Aldrich #G7513) and 1% Pen-Strep (Sigma-Aldrich #P0781). Cells were passaged at 80% confluence and fed every 2-3 days. The siRNA screen was performed using SMARTpool siRNAs (*i.e*. a pool of four siRNAs per gene) from the Mouse Druggable Subset consisting of 2,905 genes distributed on 11 384-well plates (Dharmacon #G-014675-E2, Lot 14008). The 0.1nmol lyophilised siRNA stock plates were re-suspended to 10µM, and then further diluted to give working stock plates of 3µM siRNA, which was used for transfection.

The siRNA screen was performed in 384-well round bottom plates coated with poly-D-lysine (Sigma-Aldrich #P7886). Briefly, 2.6µl of 3µM working siRNA was pipetted into each well, followed by 3µl of the transfection mix (obtained by mixing 339.6µl Lipofectamine RNAiMAX transfection reagent [Invitrogen #56532] + 4646µl OptiMEM [ThermoFisher Scientific #31985062]) and, after 5 minutes, 30µl cell suspension (diluted at 133 cell/µl). To ensure even distribution of cells into the wells, plates were incubated undisturbed at room temperature for 15 minutes before transferring carefully to the incubator at 37°C and 10% CO2 (shown to give better differentiation). Non-targeting ON-TARGETplus pool (Dharmacon #D-001810-10-20), mouse *Pparg* ON-TARGETplus SMARTpool and mouse *L3mbtl3* ON-TARGETplus SMARTpool were used as controls on each 384-well plate. All pipetting was performed using and FXP robot and Biomek AP96 P200 pipette tips (Beckman Coulter #717252).

Differentiation was initiated 24 hours post-transfection using freshly prepared insulin-oleate media: MEM-α supplemented with 0.2% FBS, 175 nM insulin, 900 µM oleate:albumin (Sigma-Aldrich #O3008) and 1% Pen-Strep and cells were incubated for 48 hours undisturbed to allow adipocyte differentiation. Following differentiation, the cells were fixed for 20 minutes by adding 6µl 24% formaldehyde directly to medium in each well. After washing with PBS, cells were stained for 30 minutes with 42µl BODIPY 493/503 for lipids (1mg/ml in ethanol; ThermoFisher Scientific #D3922) and 42µl Hoechst 33342 for nuclei (10mg/ml; Invitrogen #H3570) and washed again in PBS.

The plates were scanned on the ImageXpress Confocal Micro (Molecular Devices) using the 10x objective and the confocal pinhole setting. Cells were imaged in two channels: the Hoechst channel to capture nuclei and the FITC channel to capture the lipid droplets. To correct for unevenness across the well, three images were taken at different z planes and a maximum projection image was compiled and used for all future analyses. Images were analysed using the Multiwavelength Cell

Scoring Application Module within the MetaXpress software (Molecular Devices). Nuclei were defined in the Hoechst channel by size and intensity with user-defined thresholds. The cytoplasm was defined using the lipid stain and was thresholded by size and intensity. Cells were then classified as differentiated if the lipid stained cytoplasm exceeded a defined minimum stained area of 100µm^2 (Figure S3A). Intra-plate normalisation was carried out for each of the three triplicates, then the median of the three values for each well was calculated. This value, termed median relative differentiation (*i.e.* values normalised to non-targeting), was used to determine hits. Cuts-off of 0.7 (*i.e.* a 30% decrease in differentiation) and 1.3 (*i.e.* a 30% increase in differentiation) were used to identify genes that decrease and increase adipocyte differentiation, respectively.

siRNA knockdowns in *Drosophila* fat body

All fly strains were maintained in Darwin Chambers (IN084-AA-LT-DA-MP) at a temperature of 25°C. and 70% humidity with 12h:12h light-dark cycles, and reared on Nutri-Fly Bloomington Formulation food medium (Genesee Scientific #66-112). The fly lines used in this study were obtained from Bloomington Drosophila Stock Centre (BDSC): Ppl-Gal4 (BDSC #58768), UAS-Peb-RNAi (BDSC #33943) UAS-Ppa-RNAi (BDSC #31357), UAS-Prosalpha6T-RNAi (BDSC #55243) and UAS-Clpx-RNAi (BDSC #57577). RNAi-mediated knockdown specifically in the fat body was achieved by driving UAS-RNAi expression using the fat body specific driver line, Ppl-GAL4. Female flies from F1 generation bearing both Gal4 and UAS-RNAi constructs were used for fat body staining. F1 female progenies obtained from the cross between the Ppl-Gal4 and the UAS-RNAi lines with isogenic w1118 (BDSC #6326) wild-type flies were used as genotypic controls.

To verify the efficiency of knockdowns, total RNA was extracted from whole bodies of flies using TRI-reagent (Sigma #93289) and 2µg of total RNA was reverse transcribed using High-capacity cDNA synthesis kit (Applied Biosystems #4374967). Real-time qPCR was performed using PowerUp SYBR Green Master Mix (ThermoFisher #A25741) with listed primers (Table S4). Each qRT-PCR reaction was performed in duplicate. Rpl32 was used as an endogenous control. The cut-off for Ct values was <35 for testing genes and <25 for Rpl32. Relative expression analysis was done using the $2^{-\Delta\Delta Ct}$ method. Six to eight independent biological replicates (containing 5 whole bodies each) per genotype were tested for each experiment.

Subcuticular fat body staining of undissected fly abdomens was adapted from Li et al. (Li et al., 2017). Briefly, 2-5 days-old female flies were anesthetized to remove the legs and wings. The fly bodies were then fixed in 4% paraformaldehyde for 20 minutes, followed by washing in phosphate-

buffer saline (PBS). The flies were then submerged thrice in liquid Nitrogen for few seconds, each time followed by thawing on ice for 1 minute. A solution of 1µg/µL of BODIPY – 493/503 (1:500 dilution, ThermoFisher, #D3922) in PBS was added to the samples and incubated under dark for 1 hour, followed by three washes with PBS. The flies were then mounted onto a glass slide by gluing the thorax on the ventral side. The samples were covered with Vectashield mounting medium (Vector Laboratories #H-1000-10) and imaged by confocal microscopy (Nikon A1R) under the FITC channel. Same confocal setting was used across all samples. Maximum intensity projection images were analyzed and quantified in ImageJ.

Manual curation of 49-gene list to identify novel regulators of late-stage adipogenesis

This analysis was performed in two steps. First, three authors carried out an independent search on GWAS (https://www.ebi.ac.uk/gwas/home), OMIM (https://www.omim.org), IMPC (https://www.mousephenotype.org) and MGI (https://www.informatics.jax.org/) for evidence linking the gene of interest with obesity, lipodystrophy or other phenotypes associated with altered fat mass or adipocyte differentiation. This was supplemented with a search on PubMed (https://pubmed.ncbi.nlm.nih.gov) using a range of keywords (adipogenesis, adipocyte differentiation, lipodystrophy, obesity, adiposity, fat mass, body mass index) plus the gene symbol. The information was then verified and summarized by two other authors, leading to the identification of 19 genes without any prior evidence for a role in adipogenesis or related processes.

Analysis of MGI and IMPC adipose tissue phenotypes

We retrieved the phenotyping data for 13,487 genes from MGI, using the batch query function (data accessed on 17 May 2023). The phenotyping terms were grouped semantically in three categories: **increased fat amount** (MP:0020411 – increased abdominal adipose tissue amount, MP:0009286 – increased abdominal fat pad weight, MP:0014142 – increased body fat mass, MP:0009288 – increased epididymal fat pad weight, MP:0006094 – increased fat cell size, MP:0009285 – increased gonadal fat pad weight, MP:0009292 – increased inguinal fat pad weight, MP:0009298 – increased mesenteric fat pad weight, MP:0005458 – increased percent body fat/body weight, MP:0009302 – increased renal fat pad weight, MP:0009304 – increased retroperitoneal fat pad weight, MP:0010934 – increased subcutaneous adipose tissue amount, MP:0010024 – increased total body fat amount, MP:0008908 – increased total fat pad weight, MP:0000008 – increased white adipose tissue amount, MP:0014145 – increased white adipose tissue mass, MP:0009121 – increased white

fat cell lipid droplet size, MP:0009130 – increased white fat cell number, MP:0009118 – increased white fat cell size, and MP:0001261 – obese), **decreased fat amount** (MP:0008853 – decreased abdominal adipose tissue amount, MP:0009287 – decreased abdominal fat pad weight, MP:0014143 – decreased body fat mass, MP:0009289 – decreased epididymal fat pad weight, MP:0009269 – decreased fat cell size, MP:0009283 – decreased gonadal fat pad weight, MP:0009293 – decreased inguinal fat pad weight, MP:0009297 – decreased mammary fat pad weight, MP:0009299 – decreased mesenteric fat pad weight, MP:0009301 – decreased parametrial fat pad weight, MP:0005459 – decreased percent body fat/body weight, MP:0009303 – decreased renal fat pad weight, MP:0009305 – decreased retroperitoneal fat pad weight, MP:0008844 – decreased subcutaneous adipose tissue amount, MP:0010025 – decreased total body fat amount, MP:0008907 – decreased total fat pad weight, MP:0001783 – decreased white adipose tissue amount, MP:0014146 – decreased white adipose tissue mass, MP:0009122 – decreased white fat cell lipid droplet size, MP:0009131 – decreased white fat cell number, MP:0009133 – decreased white fat cell size, and MP:0011174 – lipodystrophy) and **abnormal fat morphology** (MP:0000010 – abnormal abdominal fat pad morphology, MP:0005452 – abnormal adipose tissue amount, MP:0011167 – abnormal adipose tissue development, MP:0000013 – abnormal adipose tissue distribution, MP:0000003 – abnormal adipose tissue morphology, MP:0012320 – abnormal body fat mass, MP:0006319 – abnormal epididymal fat pad morphology, MP:0011168 – abnormal fat cell differentiation, MP:0009115 – abnormal fat cell morphology, MP:0005334 – abnormal fat pad morphology, MP:0005335 – abnormal gonadal fat pad morphology, MP:0030880 – abnormal infrapatellar fat pad morphology, MP:0005336 – abnormal inguinal fat pad morphology, MP:0008904 – abnormal mammary fat pad morphology, MP:0008903 – abnormal mesenteric fat pad morphology, MP:0005457 – abnormal percent body fat/body weight, MP:0005337 – abnormal retroperitoneal fat pad morphology, MP:0001781 – abnormal white adipose tissue amount, MP:0014144 – abnormal white adipose tissue mass, MP:0002970 – abnormal white adipose tissue morphology, MP:0011169 – abnormal white fat cell differentiation, MP:0009117 – abnormal white fat cell morphology, MP:0009132 – abnormal white fat cell size, MP:0008901 – absent epididymal fat pad, and MP:0008843 – absent subcutaneous adipose tissue) (Table S3).

In the case of IMPC, we downloaded the lists of genes related to three phenotypes of interest: increased total body fat amount (331 genes with significant changes out of 7,575 genes tested), decreased total body fat amount (261 genes with significant changes out of 7,324 genes tested), and abnormal adipose tissue morphology (586 genes with significant changes out of 7,577 genes tested – Table S3) (data release version 18.0; data accessed on 27 March 2023). Enrichment of genes with significant phenotypes within the nine clusters of DInt was calculated using Fisher's exact tests and

represented as odds ratios ± 95% confidence intervals. To verify the specificity of our findings, we also performed a similar analysis for two unrelated phenotypes (namely, abnormal heart electrocardiography waveform feature and abnormal fear/anxiety−related behaviour), neither of which associated any significant enrichment in the nine clusters of DInt genes.

Enrichment of adiposity-related GWAS genes in the nine clusters of DInt

Human adiposity-related GWAS SNPs and their reported target genes (locus genes for adipose storage capacity) were downloaded from the GWAS catalogue (obesity:EFO_0001073; adipose tissue measurement:EFO_0004764). Fisher's tests were applied to check for enrichment of these sets of genes within the gene lists found in each of the nine clusters.

Mouse-human synteny analyses

We have assessed synteny based on synteny blocks identified using the synteny portal (Lee et al., 2016), and defined the exact homologous positions from the UCSC hg19-mm9 syntenic Net files. Differential interactions, where both fragments fall into the same synteny block, and obesity-related intergenic, intronic and TF binding site SNPs falling into synteny blocks were identified. Those SNPs that overlapped with the distal fragment of a differential interaction were mapped to the gene involved in that interaction.

TF enrichment analysis

Up- and down-regulated CAGE-seq peaks were overlapped with the promoter and distal fragments of differential interactions across the nine clusters. The positions of CAGE-seq peaks (>7bp) were used in i-cisTarget (Imrichova et al., 2015) to test for enrichment of known motifs and TF binding based on publicly available ChIP-seq data. TFs with a normalized enrichment score (NES) above 5 were considered significant.

Pathway analysis

Genes involved in differential interactions across the nine clusters were tested for enrichment of GO biological processes using g:Profiler (Raudvere et al., 2019). The lists of significantly enriched (p-adj <0.05) GO terms were summarized and visualized with REVIGO (Supek et al., 2011).

Candidate genes GWAS and eQTL analyses

The novel 19 genes encoding regulators of late adipogenesis were integrated with genome-wide association studies (GWAS) and related functional annotations, pertaining to expressions quantitative trait loci (eQTL) datasets and nearby enhancer elements. For this, we used GWAS data on body mass index (BMI) and waist-hip ratio (WHR) adjusted for BMI in up to 806,834 and 694,649 individuals, respectively, available from the UK Biobank and the GIANT consortium (Pulit et al., 2019). From these GWAS, we looked at common variants with a minor allele frequency >0.1%.

For each of the target genes and each phenotypic trait, genes were annotated based on proximity to genome-wide significant signals ($p<5x10^{-8}$), in 1Mb windows; 500kb up- or downstream of the genes start or end site. For genes with proximal GWAS signals, we calculated genomic windows of high linkage disequilibrium (LD; $R^2>0.8$) for each given signal and mapped these to the locations of known enhancers for the target genes, using the activity-by-contact (ABC) enhancer maps (Nasser et al., 2021). For genomic variants reaching at least a suggestive level of significance in the GWAS ($p<5x10^{-5}$), we performed SMR and HEIDI tests (v1.02, Zhu et al., 2016) using blood gene expression level data from the eQTLGen study (Võsa et al., 2021) and the GTEx cross-tissue meta-analysis (V7, GTEx Consortium, 2015 available via https://gtexportal.org and using the fixed-effects summary statistics). For the eQTL analyses, we considered gene expression of a gene to be influenced by the same genomic variation as that seen in the GWAS, if the FDR-corrected p-value for the SMR-test was p<0.05 and the p-value for the HEIDI test was >0.1%.

Statistical analyses

Statistical analyses was performed as described above or using GraphPad Prism 9 software. For all tests, P<0.05 was considered significant.

Data availability

Sequencing data w deposited in NCBI's Gene Expression Omnibus (GEO) under the accession numbers GEO: GSE234744 (RNA-seq of primary pre-adipocytes and adipocytes isolated from 3 months-old hemizygous Zfp423^{GFP} mice), GSE234747 (PCHi-C-seq of D0 and D7 3T3-L1 cells) and GSE234749 (CAGE-seq of D0 and D7 3T3-L1 cells).

Acknowledgments

This work was supported by the Medical Research Council (MR/J001562/1 to M.C.; MRC_MC_UU_12014/4 to M.C. and S.E.O.; MRC_MC_UU_12012/5 to the MRC Metabolic Diseases Unit; MRC_MC_UU_12015/2 and MRC_MC_UU_00006/2 to J.R.B.P.) and the Wellcome Trust (214274/Z/18/Z to S.O.R). This research was also conducted using the UK Biobank Resource under application 9905. S.S. was supported by a UKRI MRC Rutherford Fund Fellowship (MR/T016787/1), a BBSRC Institute Strategic Programme grant (BBS/E/B/000C0421), and a Career Progression Fellowship from the Babraham Institute. F.M. was supported by the Qatar National Research Fund (UREP28-269-1-051 and NPRP14S-0319-210075). A.B. and F.M. were supported by funding from CHLS (College of Health & Life Sciences, Hamad Bin Khalifa University, Doha, Qatar). A.E., B.S.H. and A.T.C. were funded by MRC Programme Grants MC_UU_12022/1 and MC_UU_12022/8 awarded to A.R.V.

Author contributions

M.C. and S.E.O. designed the project; M.C., S.E.O., F.M., S.O.R., J.R.B.P. and A.R.V. secured the funding; I.S., B.M., A.E., P.G., K.A.K., P.F., S.S., F.M., S.O.R., J.R.B.P., A.R.V., S.E.O. and M.C. designed the experimental setup; I.S., A.E., P.G., A.B., N.C., B.S.H., A.T.C., D.S.F.-T., S.S. and F.M. performed all in vitro and in vivo experimental work; B.M., I.S., L.S., A.E., P.G., K.A.K., A.B., D.S.F.-T., L.V.M., L.S., S.A., D.C., R.S.H., S.W.W., K.K.O., S.S., F.M. and J.R.B.P. performed data analysis and interpretation; I.S., B.M., S.E.O. and M.C. wrote the paper. All other authors discussed the results and edited the manuscript. M.C. managed and supervised all aspects of the study.

Declaration of interests

S.S. is a co-founder and shareholder of Enhanc3D Genomics. J.R.B.P. is an employee and shareholder of Adrestia Therapeutics Ltd. and receives research funding from GSK.

References

1. Agrawal, S., Klarqvist, M.D.R., Diamant, N., Stanley, T.L., Ellinor, P.T., Mehta, N.N., Philippakis, A., Ng, K., Claussnitzer, M., Grinspoon, S.K., et al. (2023). BMI-adjusted adipose tissue volumes exhibit depot-specific and divergent associations with cardiometabolic diseases. Nat. Commun. *14*, 266.

2. Agrawal, S., Wang, M., Klarqvist, M.D.R., Smith, K., Shin, J., Dashti, H., Diamant, N., Choi, S.H., Jurgens, S.J., Ellinor, P.T., et al. (2022). Inherited basis of visceral, abdominal subcutaneous and gluteofemoral fat depots. Nat. Commun. *13*, 3771.
3. Akbari, P., Sosina, O.A., Bovijn, J., Landheer, K., Nielsen, J.B., Kim, M., Aykul, S., De, T., Haas, M.E., Hindy, G., et al. (2022). Multiancestry exome sequencing reveals INHBE mutations associated with favorable fat distribution and protection from diabetes. Nat. Commun. *13*, 4844.
4. Bahrami-Nejad, Z., Zhang, Z.B., Tholen, S., Sharma, S., Rabiee, A., Zhao, M.L., Kraemer, F.B., and Teruel, M.N. (2022). Early enforcement of cell identity by a functional component of the terminally differentiated state. PLoS Biol. *20*, e3001900.
5. Berry, R., and Rodeheffer, M.S. (2013). Characterization of the adipocyte cellular lineage in vivo. Nat. Cell Biol. *15*, 302–308.
6. Chang, E.J., Ha, J., Kang, S.S., Lee, Z.H., and Kim, H.H. (2011). AWP1 binds to tumor necrosis factor receptor-associated factor 2 (TRAF2) and is involved in TRAF2-mediated nuclear factor-kappaB signaling. Int. J. Biochem. Cell Biol. *43*, 1612–1620.
7. Choy, M.K., Javierre, B.M., Williams, S.G., Baross, S.L., Liu, Y., Wingett, S.W., Akbarov, A., Wallace, C., Freire-Pritchett, P., Rugg-Gunn, P.J., et al. (2018). Promoter interactome of human embryonic stem cell-derived cardiomyocytes connects GWAS regions to cardiac gene networks. Nat. Commun. *9*, 2526.
8. Chu, A.Y., Deng, X., Fisher, V.A., Drong, A., Zhang, Y., Feitosa, M.F., Liu, C.T., Weeks, O., Choh, A.C., Duan, Q., et al. (2017). Multiethnic genome-wide meta-analysis of ectopic fat depots identifies loci associated with adipocyte development and differentiation. Nat. Genet. *49*, 125-130.
9. Colombani, J., Raisin, S., Pantalacci, S., Radimerski, T., Montagne, J., and Léopold, P. (2003). A nutrient sensor mechanism controls Drosophila growth. Cell *114*, 739–749.
10. Dekker, J. (2006). The three 'C's of chromosome conformation capture: controls, controls, controls. Nat Methods. *3*, 17–21.
11. Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.
12. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
13. Edupuganti, R.R., Harikumar, A., Aaronson, Y., Biran, A., Sailaja, B.S., Nissim-Rafinia, M., Azad, G.K., Cohen, M.A., Park, J.E., Shivalila, C.S., et al. (2017). Alternative SET/TAFI Promoters Regulate Embryonic Stem Cell Differentiation. Stem Cell Reports *9*, 1291–1303.

14. Feng, G., Tong, M., Xia, B., Luo, G.Z., Wang, M., Xie, D., Wan, H., Zhang, Y., Zhou, Q., and Wang, X.J. (2016). Ubiquitously expressed genes participate in cell-specific functions via alternative promoter usage. EMBO Rep. *17*, 1304–1313.
15. Ghaben, A.L., and Scherer, P.E. (2019). Adipogenesis and metabolic health. Nat. Rev. Mol. Cell. Biol. *20*, 242–258.
16. Green, H. and Kehinde, O. (1974). Sublines of mouse 3T3 cells that accumulate lipid. Cell *1*, 113–116.
17. Green, H., and Meuth, M. (1974). An established pre-adipose cell line and its differentiation in culture. Cell *3*, 127–133.
18. GTEx Consortium. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.
19. Gupta, R.K., Mepani, R.J., Kleiner, S., Lo, J.C., Khandekar, M.J., Cohen, P., Frontini, A., Bhowmick, D.C., Ye, L., Cinti, S., et al. (2012). Zfp423 expression identifies committed preadipocytes and localizes to adipose endothelial and perivascular cells. Cell Metab. *15*, 230–239.
20. Haak, M., Vinke, S., Keller, W., Droste, J., Rückert, C., Kalinowski, J., and Pucker, B. (2018). High Quality de Novo Transcriptome Assembly of Croton tiglium. Front. Mol. Biosci. *5*, 62.
21. Hartigan, J.A., and Wong, M.A. (1979). A K-Means Clustering Algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. *28*, 100–108.
22. He, M., Li, Y., Tang, Q., Li, D., Jin, L., Tian, S., Che, T., He, S., Deng, L., Gao, G., et al. (2018). Genome-Wide Chromatin Structure Changes During Adipogenesis and Myogenesis. Int. J. Biol. Sci. *14*, 1571–1585.
23. Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Mägi, R., et al. (2010). Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat. Genet. *42*, 949–960.
24. Huang, L.O., Rauch, A., Mazzaferro, E., Preuss, M., Carobbio, S., Bayrak, C.S., Chami, N., Wang, Z., Schick, U.M., Yang, N., et al. (2021). Genome-wide discovery of genetic loci that uncouple excess adiposity from its comorbidities. Nat. Metab. *3*, 228–243.
25. Imrichová, H., Hulselmans, G., Atak, Z.K., Potier, D., and Aerts, S. (2015). i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. Nucleic Acids Res. *43*, W57–W64.
26. James, D.E., Stöckli, J., and Birnbaum, M.J. (2021). The aetiology and molecular landscape of insulin resistance. Nat. Rev. Mol. Cell. Biol. *22*, 751–771.

27. Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell *167*, 1369–1384.
28. Jiang, S., Wei, H., Song, T., Yang, Y., Zhang, F., Zhou, Y., Peng, J., and Jiang, S. (2015). KLF13 promotes porcine adipocyte differentiation through PPARγ activation. Cell Biosci. *5*, 28.
29. Kang, S., Akerblad, P., Kiviranta, R., Gupta, R.K., Kajimura, S., Griffin, M.J., Min, J., Baron, R., and Rosen, E.D. (2012). Regulation of early adipose commitment by Zfp521. PLoS Biol. *10*, e1001433.
30. Kim, H., Westerman, K.E., Smith, K., Chiou, J., Cole, J.B., Majarian, T., von Grotthuss, M., Kwak, S.H., Kim, J., Mercader, J.M., et al. (2023). High-throughput genetic clustering of type 2 diabetes loci reveals heterogeneous mechanistic pathways of metabolic disease. Diabetologia *66*, 495–507.
31. Lane, J.M., Doyle, J.R., Fortin, J.P., Kopin, A.S., and Ordovás, J.M. (2014). Development of an OP9 derived cell line as a robust model to rapidly study adipocyte differentiation. PLoS One. *9*, e112123.
32. Lee, J., Hong, W.Y., Cho, M., Sim, M., Lee, D., Ko, Y., and Kim, J. (2016). Synteny Portal: a web-based application portal for synteny block analysis. Nucleic Acids Res. *44*, W35–W40.
33. Li, D., Yea, S., Li, S., Chen, Z., Narla, G., Banck, M., Laborda, J., Tan, S., Friedman, J.M., Friedman, S.L., et al. (2005). Krüppel-like factor-6 promotes preadipocyte differentiation through histone deacetylase 3-dependent repression of DLK1. J. Biol. Chem. *280*, 26941–26952.
34. Li, Y., Tiedemann, L., von Frieling, J., Nolte, S., El-Kholy, S., Stephano, F., Gelhaus, C., Bruchhaus, I., Fink, C., and Roeder, T. (2017). The Role of Monoaminergic Neurotransmission for Metabolic Control in the Fruit Fly Drosophila Melanogaster. Front. Syst. Neurosci. *11*, 60.
35. Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108.
36. Link, J.C., Wiese, C.B., Chen, X., Avetisyan, R., Ronquillo, E., Ma, F., Guo, X., Yao, J., Allison, M., Chen, Y.I., et al. (2020). X chromosome dosage of histone demethylase KDM5C determines sex differences in adiposity. J. Clin. Invest. *130*, 5688–5702.
37. Loh NY, Neville MJ, Marinou K, Hardcastle SA, Fielding BA, Duncan EL, McCarthy MI, Tobias JH, Gregson CL, Karpe F, et al. (2015). LRP5 regulates human body fat distribution by modulating adipose progenitor biology in a dose- and depot-specific fashion. Cell Metab. *21*, 262–273.
38. Lotta, L.A., Gulati, P., Day, F.R., Payne, F., Ongen, H., van de Bunt, M., Gaulton, K.J., Eicher, J.D., Sharp, S.J., Luan, J., et al. (2017). Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. Nat. Genet. *49*, 17–26.

39. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
40. Mann, J.P., and Savage, D.B. (2019). What lipodystrophies teach us about the metabolic syndrome. J. Clin. Invest. *129*, 4009–4021.
41. Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y., Kimura, H., Tanaka, T., Tsutsumi, S., Nakao, M., Doi, T., et al. (2015). H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. Mol. Cell. *60*, 584–596.
42. Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat. Genet. *47*, 598–606.
43. Mikkelsen, T.S., Xu, Z., Zhang, X., Wang, L., Gimble, J.M., Lander, E.S., and Rosen, E.D. (2010). Comparative epigenomic analysis of murine and human adipogenesis. Cell. *143*, 156–169.
44. Molyneaux, B.J., Goff, L.A., Brettler, A.C., Chen, H.H., Hrvatin, S., Rinn, J.L., and Arlotta, P. (2015). DeCoN: genome-wide analysis of *in vivo* transcriptional dynamics during pyramidal neuron fate selection in neocortex. Neuron *85*, 275–288.
45. Murata, M., Nishiyori-Sueki, H., Kojima-Ishiyama, M., Carninci, P., Hayashizaki, Y., and Itoh, M. (2014). Detecting expressed genes using CAGE. Methods Mol. Biol. 1164, 67–85.
46. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, Jones TR, Nguyen TH, Ulirsch JC, Lekschas F, et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. Nature *593*, 238–243.
47. Nicetto, D., and Zaret, K.S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Curr. Opin. Genet. Dev. *55*, 1–10.
48. Nichols, M.H., and Corces, V.G. (2021). Principles of 3D compartmentalization of the human genome. Cell Rep. *35*, 109330.
49. Ntambi, J.M. and Young-Cheul, K. (2000). Adipocyte differentiation and gene expression. J. Nutr. *130*, 3122S–3126S.
50. Oh, R., Heo, J.W., Kim, H., Mi-Kyung Sung, M.K., and Kim, S.E. (2021). Effects of Sex-Specific Gene on the Adipogenic Differentiation in 3T3-L1 Preadipocytes. Curr. Dev. Nutr. *5*,1237.
51. Ohmiya, H., Vitezic, M., Frith, M.C., Itoh, M., Carninci, P., Forrest, A.R., Hayashizaki, Y., Lassmann, T., and FANTOM Consortium. (2014). RECLU: a pipeline to discover reproducible transcriptional start sites and their alternative regulation using capped analysis of gene expression (CAGE). BMC Genomics. *15*, 269.
52. Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat. Rev. Genet. *12*, 283–293.

53. Park, J., Lee, D.H., Ham, S., Oh, J., Noh, J.R., Lee, Y.K., Park, Y.J., Lee, G., Han, S.M., Han, J.S., et al. (2022). Targeted erasure of DNA methylation by TET3 drives adipogenic reprogramming and differentiation. Nat. Metab. *4*, 918–931.
54. Peters, L.L., Robledo, R.F., Bult, C.J., Churchill, G.A., Paigen, B.J., and Svenson, K.L. (2007). The mouse as a model for human biology: a resource guide for complex trait analysis. Nat. Rev. Genet. *8*, 58–69.
55. Pospisilik, J.A., Schramek, D., Schnidar, H., Cronin, S.J., Nehme, N.T., Zhang, X., Knauf, C., Cani, P.D., Aumayr, K., Todoric, J., et al. (2010). Drosophila genome-wide obesity screen reveals hedgehog as a determinant of brown versus white adipose cell fate. Cell. *140*, 148–160.
56. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum. Mol. Genet. *28*, 166–174.
57. Ramírez-Zacarías, J.L., Castro-Muñozledo, F., and Kuri-Harcuch, W. (1992). Quantitation of adipose conversion and triglycerides by staining intracytoplasmic lipids with Oil red O. Histochemistry *97*, 493–497.
58. Rask-Andersen, M., Karlsson, T., Ek, W.E., and Johansson, Å. (2019). Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. Nat. Commun. *10*, 339.
59. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. *47*, W191–W198.
60. Rosen, E.D., and Spiegelman, B.M. (2014). What we talk about when we talk about fat. Cell *156*, 20–44.
61. Rubin, A.J., Barajas, B.C., Furlan-Magaril, M., Lopez-Pajares, V., Mumbach, M.R., Howard, I., Kim, D.S., Boxer, L.D., Cairns, J., Spivakov, M., et al. (2017). Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. Nat. Genet. 49, 1522–1528.
62. Schleinitz, D., Böttcher, Y., Blüher, M., Kovacs, P. (2014). The genetics of fat distribution. Diabetologia *57*, 1276–1286.
63. Schoenfelder, S., Mifsud, B., Senner, C.E., Todd, C.D., Chrysanthou, S., Darbo, E., Hemberger, M., and Branco, M.R. (2018). Divergent wiring of repressive and active chromatin interactions between mouse embryonic and trophoblast lineages. Nat. Commun. *9*, 4189.

64. Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. Nat. Rev. Genet. *20*, 437–455.
65. Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., et al. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. *25*, 582–597.
66. Siersbæk, R., Madsen, J.G.S., Javierre, B.M., Nielsen, R., Bagge, E.K., Cairns, J., Wingett, S.W., Traynor, S., Spivakov, M., Fraser, P., et al. (2017). Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. Mol. Cell. *66*, 420–435.
67. Song, M., Zheng, Y., Qi, L., Hu, F.B., Chan, A.T., and Giovannucci, E.L. (2018). Longitudinal Analysis of Genetic Susceptibility and BMI Throughout Adult Life. Diabetes *67*, 248–255.
68. Stefkovich, M., Traynor, S., Cheng, L., Merrick, D., and Seale, P. (2021). Dpp4+ interstitial progenitor cells contribute to basal and high fat diet-induced adipogenesis. Mol. Metab. *54*, 101357.
69. Supek, F., Bošnjak, M., Škunca, N., and Šmuc T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800.
70. Vacca, A., Itoh, M., Kawaji, H., Arner, E., Lassmann, T., Daub, C.O., Carninci, P., Forrest, A.R.R., Hayashizaki, Y., FANTOM Consortium, et al. (2018). Conserved temporal ordering of promoter activation implicates common mechanisms governing the immediate early response across cell types and stimuli. Open Biol. *8*, 180011.
71. Võsa, U., Claringbould, A., Westra, H.J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Yazar, S., et al. (2021). Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat. Genet. *53*, 1300–1310.
72. Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. F1000Res. *4*, 1310.
73. Winkler, T.W., Justice, A.E., Graff, M., Barata, L., Feitosa, M.F., Chu, S., Czajkowski, J., Esko, T., Fall, T., Kilpeläinen, T.O., et al. (2015). The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. PLoS Genet. *11*, e1005378.
74. Wolins, N.E., Quaynor, B.K., Skinner, J.R., Tzekov, A., Park, C., Choi, K., and Bickel, P.E. (2006). OP9 mouse stromal cells rapidly differentiate into adipocytes: characterization of a useful new model of adipogenesis. J. Lipid Res. *47*, 450–460.

75. Wragg, J.W., Roos, L., Vucenovic, D., Cvetesic, N., Lenhard, B., and Müller, F. (2020). Embryonic tissue differentiation is characterized by transitions in cell cycle dynamic-associated core promoter regulation. Nucleic Acids Res. *48*, 8374–8392.
76. Yamakawa, D., Katoh, D., Kasahara, K., Shiromizu T., Matsuyama, M., Matsuda, C., Maeno, Y., Watanabe, M., Nishimura, Y., and Inagaki, M. (2021). Primary cilia-dependent lipid raft/caveolin dynamics regulate adipogenesis. Cell Rep. *34*, 108817.
77. Zhang, N., Mendieta-Esteban, J., Magli, A., Lilja, K.C., Perlingeiro, R.C.R., Marti-Renom, M.A., Tsirigos, A., and Dynlacht, B.D. (2020). Muscle progenitor specification and myogenic differentiation are associated with changes in chromatin topology. Nat. Commun. *11*, 6222.
78. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.
79. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., et al. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487.

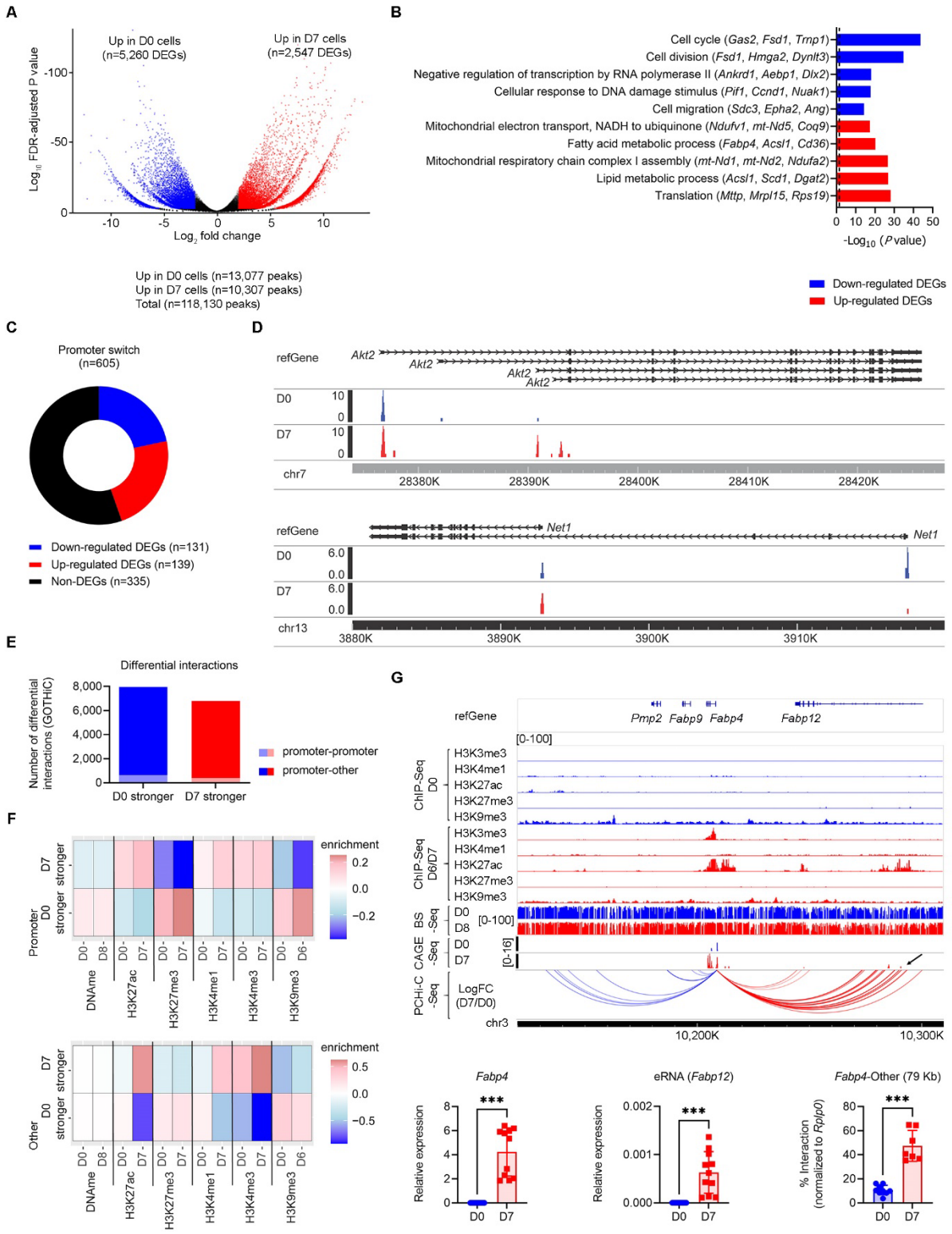

Figure 1. Transcriptional and epigenetic changes identified in the 3T3-L1 model of *in vitro* adipogenesis

(A) Volcano plot depicting differentially expressed peaks (*i.e.* transcripts) between D0 and D7 3T3-L1 cells, as identified by CAGE-seq. Approximately 16.5% of all transcripts (23.3K) that belong to 7,807 genes are differentially expressed as result of *in vitro* adipogenesis.

(B) DAVID analysis depicting biological processes enriched in differentially expressed genes (overall fold change >2). Dotted line corresponds to FDR-corrected P value <0.05.

(C) Donut chart showing the distribution of the alternative usage of transcripts relative to the overall level of gene expression for the 605 genes that show promoter switching during 3T3-L1 adipogenesis.

(D) Examples of two genes that show a promoter switching (blue – peaks that are expressed at higher levels at D0, red – peaks that are expressed at higher level at D7).

(E) Differential D0/D7 promoter-anchored interactions identified by promoter-capture Hi-C and quantified using GOTHiC. Only a minority of these differential interactions engage other promoters (promoter-promoter interactions), while the most are with intergenic or intragenic regions (promoter-other).

(F) Epigenetic signatures at the promoter area (top panel) and distal interacting fragment (other – bottom panel) that are engaged in differential D0/D7 interactions.

(G) Top: example of a differential D0/D7 interaction engaging *Fabp4* promoter with an enhancer-like region located 79 kb downstream, within one of the introns of *Fabp12* locus (see arrow, marked at D7 by increased levels of H3K27ac and active transcription – CAGE-seq peaks). Interactions that are stronger and weaker at D7 are shown in red and blue, respectively. Bottom: Validation of *Fabp4* mRNA changes and eRNA expression (by qRT-PCR) and *Fabp4*-Other interaction (by q3C) in independent biological replicates. Error bars are standard deviation (SD), *** P<0.001 by Mann-Whitney tests.

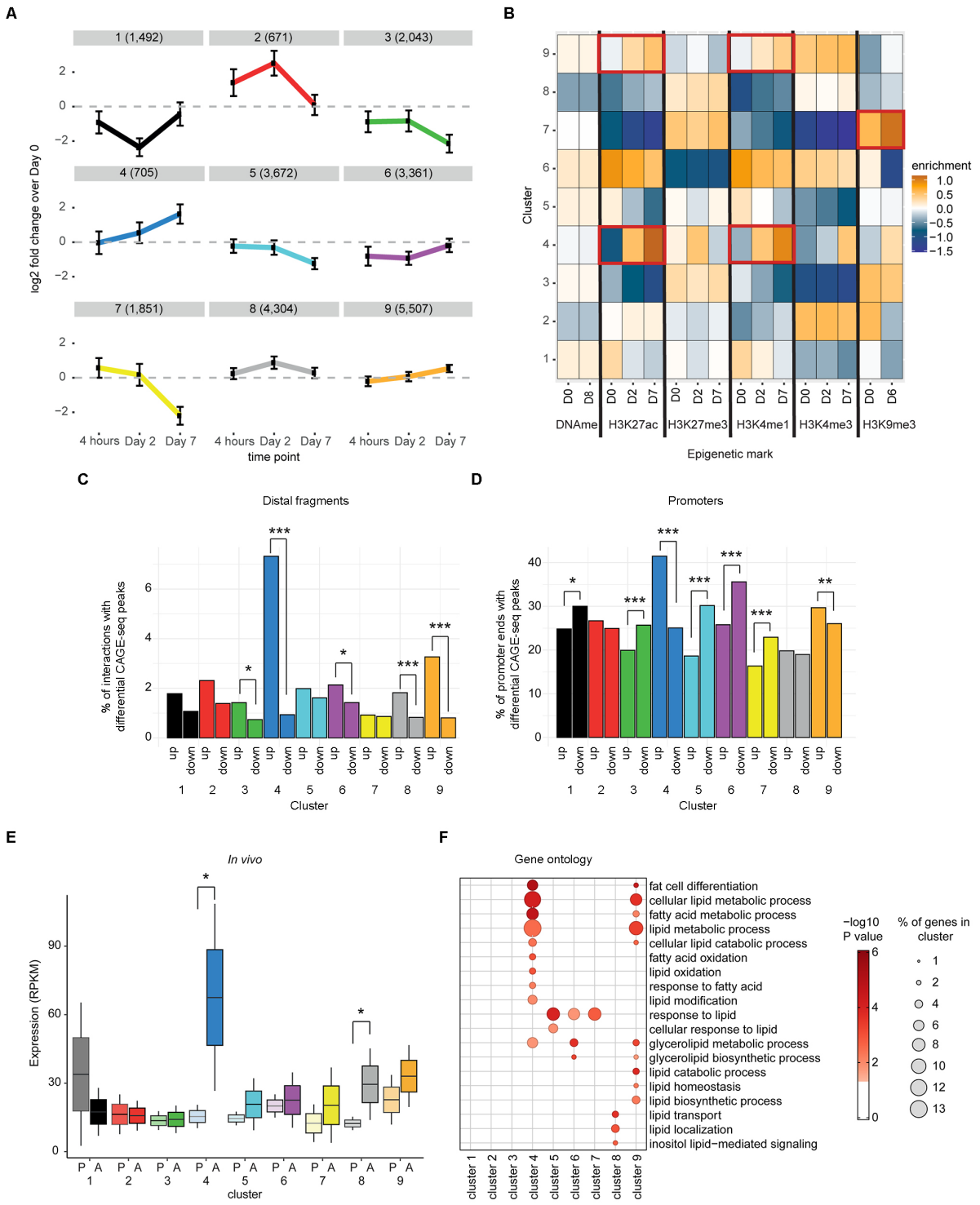

Figure 2. Cluster analysis of differential promoter-anchored interactions

(A) The nine clusters of differential interactions (DInt) identified based on their dynamics during D0, 4h, D2 and D7 stages of 3T3-L1 adipogenesis (numbers shown between parentheses indicate number of DInt/cluster). Interaction frequencies are shown as Log2 fold changes relative to D0 (horizontal dotted lines). Clusters 3, 4, 5, 7 and 9 are associated with significant changes at the transition D2 – D7 of adipogenesis. Error bars represent SD.

(B) Dynamics of epigenetic marks at the distal regions (other) engaged in DInt. Cluster 4 shows significant increases in H3K27ac and H3K4me1 (marking active enhancers) at the transition D2-D7, while cluster 7 associates with a marked increase of the repressive mark H3K9me3. Enrichment is relative to the average level of the modification across the nine clusters.

(C) Percentages of distal fragments (other) per cluster containing differential CAGE-seq peaks. Clusters 3, 4, 6, 8 and 9 DInt are enriched in CAGE-seq peaks upregulated at D7, suggestive of enhancer activity. * P<0.05, *** P<0.001 by Fisher's exact tests.

(D) Percentages of promoters per cluster containing differential CAGE-seq peaks. Clusters 4 and 9 are enriched in CAGE-seq peaks up-regulated during adipocyte differentiation, while clusters 1, 3, 5, 6 and 7 are enriched in CAGE-seq peaks down-regulated during adipocyte differentiation. * P<0.05, *** P<0.001 by Fisher's exact tests.

(E) Genes that belong to clusters 4 and 8 have a higher mean level of expression in adipocytes versus pre-adipocytes *in vivo*, suggesting potential enrichment in regulators of adipogenesis. Data is shown as box plots with mean (horizontal line) and 95% confidence intervals (error bars). * P<0.05 by paired t tests. P: preadipocyte, A: adipocyte.

(F) Dot plot depicting GO terms related to adipocyte differentiation and lipid metabolism enriched in genes that are part of DInt clusters. The colour and the size of each dot indicate the strength of significance and the percentage of genes in each cluster, respectively.

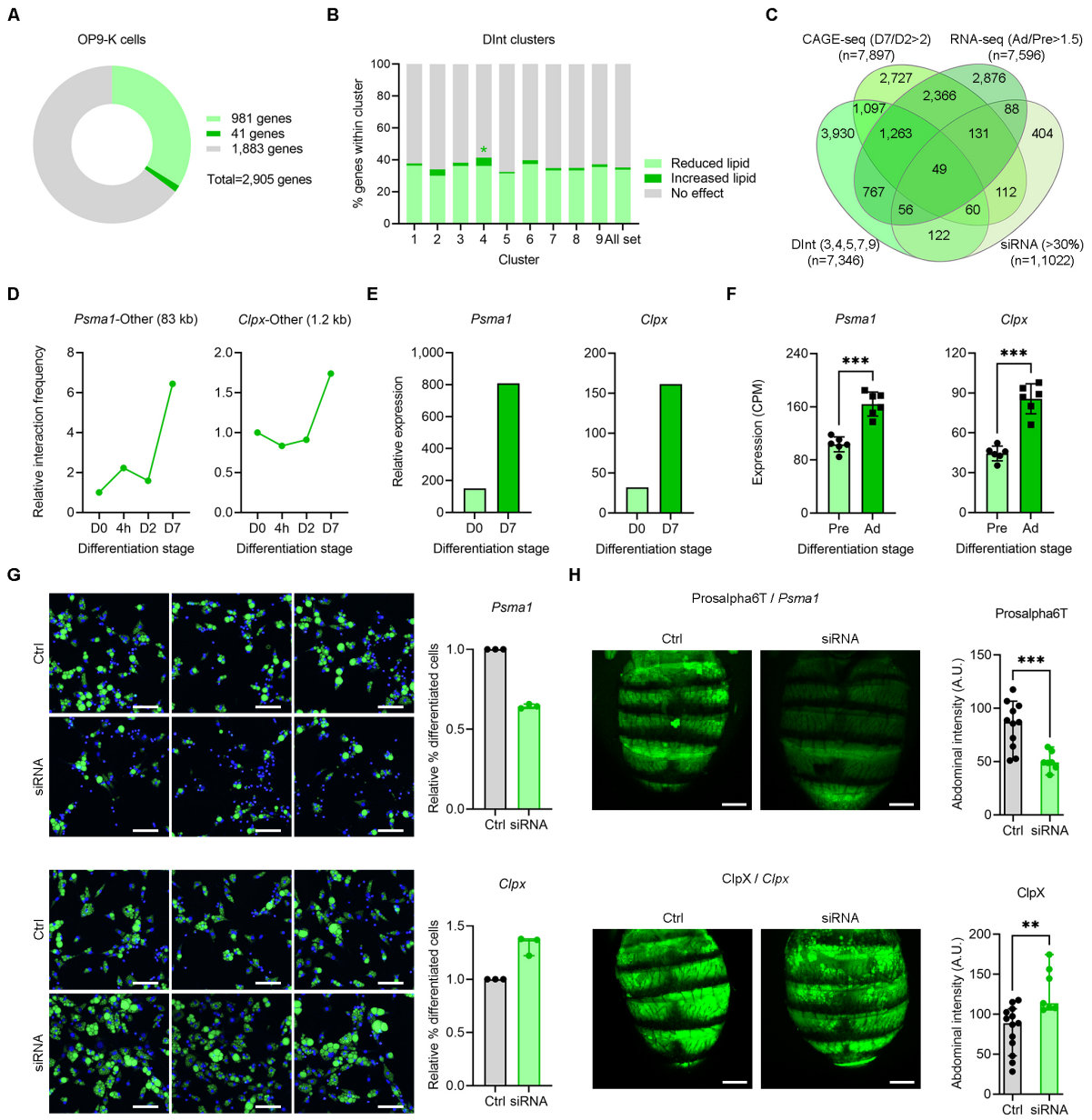

Figure 3. Validation of novel regulators of late-stage adipogenesis by siRNA knockdowns *in vitro* and *in vivo*

(A) Overview of the results of a siRNA screen in OP9-K cells: light green – >30% reduction in lipid accumulation, dark green – >30% increase in lipid accumulation, grey – less than 30% change in lipid accumulation.

(B) Proportions of genes, per cluster, that affect lipid accumulation in OP9-K cells upon siRNA knockdown. Cluster 4 is significantly enriched in genes that increase lipid droplet accumulation upon siRNA knockdown (* P<0.05 by Fisher's exact test).

(C) Venn diagram analysis identifies 49 candidates of late adipogenesis that fulfil the four criteria (see main text).

(D) Dynamics of *Psma1* and *Clpx* DInt during 3T3-L1 adipocyte differentiation as quantified by GOTHiC.

(E) Relative expression of *Psma1* and *Clpx* in 3T3-L1 pre-adipocytes (D0) and adipocytes (D7) measured by CAGE-seq.

(F) Relative expression of *Psma1* and *Clpx* in primary mouse pre-adipocytes and adipocytes measured by RNA-seq (CPM – counts per million). Error bars are SD (standard deviation), *** P<0.001.

(G) Left: representative images of OP9-K adipocytes stained with Bodipy (green, marking lipid droplets) + DAPI (blue, marking nuclei), upon siRNA knockdown of *Psma1* and *Clpx* (scale bars are 50 µm). Right: quantification of % of differentiated cells, shown relative to controls (Ctrl), arbitrarily normalized to 1. Error bars are SD.

(H) Left: representative images of Bodipy-stained *Drosophila* abdomens upon Prosalpha6T and ClpX siRNA treatment (scale bars are 200 µm). Right: quantification of Bodipy-staining intensity in the fat body. Error bars are SD, ** P<0.01, *** P<0.001 by unpaired t tests with Welch's correction.

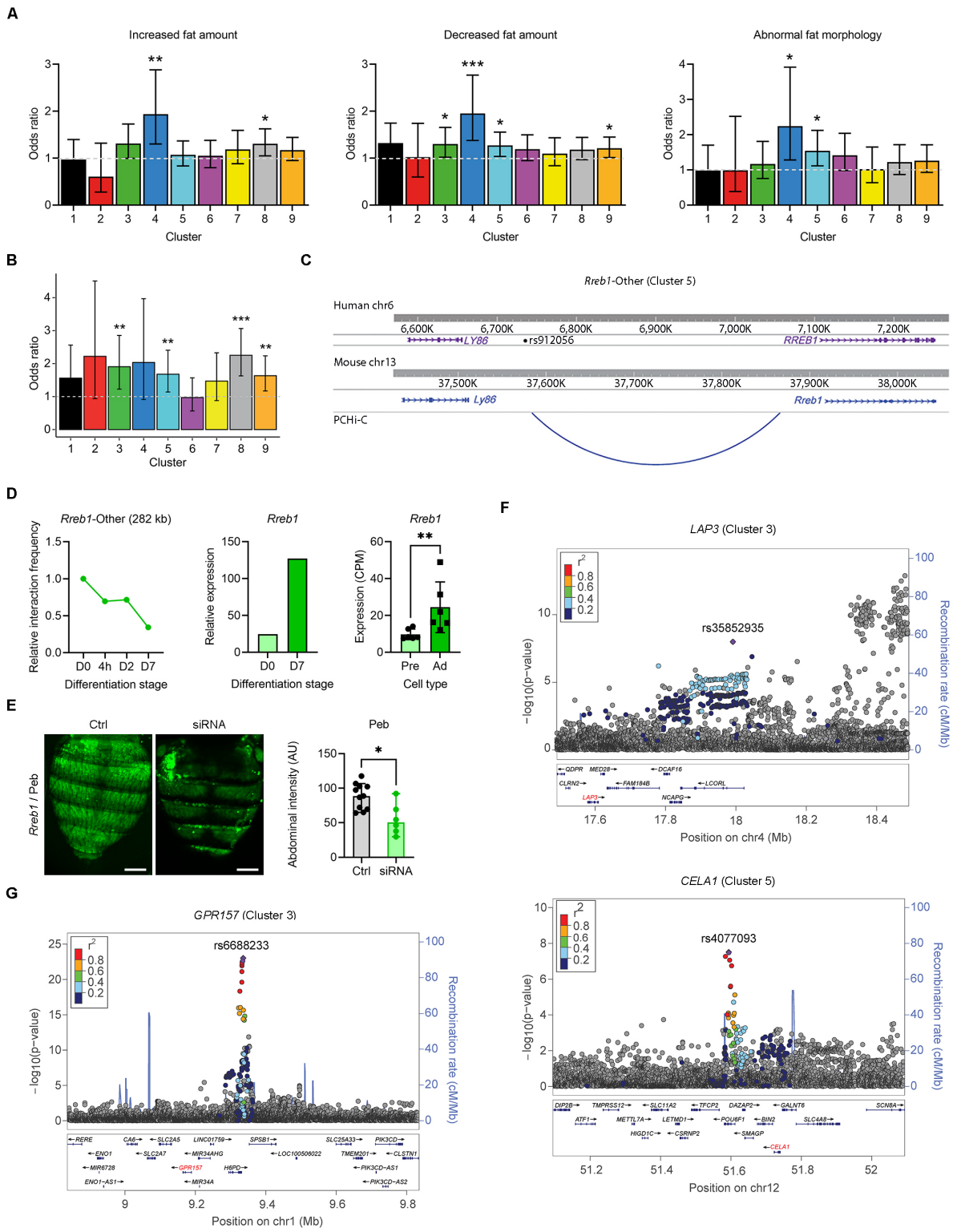

Figure 4. Validation of novel regulators of late-stage adipogenesis using mouse and human genetic analyses

(A) MGI (Mouse Genome Informatics) data showing significant enrichment of genes per DInt cluster, whose mutations lead to adipose tissue-related phenotypes *in vivo*. Columns indicate odds ratios

and error bars are 95% confidence intervals. * P<0.05, ** P<0.01, *** P<0.001 by Fisher's exact tests.

(B) Clusters that are enriched in GWAS candidate genes associated with human adiposity related traits. Columns indicate odds ratios and error bars are 95% confidence intervals. * P<0.05, ** P<0.01, *** P<0.001 by Fisher's exact tests.

(C) Synteny analysis between a region on human chromosome 6 (containing rs912056 linked to fat distribution traits by GWAS and implicating *LY86* as causative gene) and the homologous region on mouse chromosome 13 (showing a cluster 5 DInt that becomes weaker at D7 – blue colour – and links the promoter of *Rreb1*, instead of *Ly86*, with the region syntenic with the rs912056 location).

(D) Left: quantification of *Rreb1*-other DInt by GOTHiC, shown relative to D0. Middle: *Rreb1* up-regulation during *in vitro* adipogenesis in 3T3-L1 cells (as quantified by CAGE-seq). Right: *Rreb1* up-regulation during *in vivo* adipogenesis (as measured by RNA-seq). Error bars are SD, ** P<0.01.

(E) siRNA knockdown of Peb (the *Drosophila* homologue gene of *Rreb1*) leads to reduced lipid accumulation in the fat body. Left: representative images of Bodipy-stained *Drosophila* abdomens upon siRNA treatment (scale bars are 200 µm). Right: quantification of BODIPY-staining intensity in the fat body. Error bars are SD, * P<0.05 by an unpaired t test with Welch's correction.

(F) Regional association plot of GWAS hits near *LAP3* (rs35852935) and *GPR157* (rs6688233) identified as regulators of BMI.

(G) Regional association plot of a GWAS hit near *GPR157* (rs6688233) identified as regulator of BMI-adjusted WHR (waist-hip ratio). For panels (F) and (G) the strength of association (r^2) is colour-coded. The genetic variant and the associated gene are in linkage disequilibrium, as indicated by the recombination rates.