3D genomic features across >50 diverse cell types reveal insights into the genomic architecture of childhood obesity

Khanh B. Trang, PhD[1,2] ; Matthew C. Pahl, PhD[1,2]; James A. Pippin, BA[1,2]; Chun Su, PhD[1,2]; Sheridan H. Littleton, BS[1,2,3,4]; Prabhat Sharma, PhD[1,5]; Nikhil N. Kulkarni, PhD[1,5]; Louis R. Ghanem, MD, PhD[6]; Natalie A. Terry, MD, PhD[6]; Joan M. O'Brien, MD[7,8]; Yadav Wagley, PhD[9]; Kurt D. Hankenson, DVM, MS, PhD[9]; Ashley Jermusyk, PhD[10]; Jason W. Hoskins, PhD[10]; Laufey T. Amundadottir, PhD[10]; Mai Xu, PhD[10]; Kevin M Brown, PhD[10]; Stewart A. Anderson, PhD[11,12]; Wenli Yang, PhD[13,14]; Paul M. Titchenell, PhD[13,15]; Patrick Seale, PhD[13,14]; Laura Cook, PhD[16,17,18]; Megan K. Levings, PhD[19,20,21]; Babette S. Zemel, PhD[6,22]; Alessandra Chesi, PhD[1,23]; Andrew D. Wells, PhD[1,5,23,24]; Struan F.A. Grant, PhD[1,2,4,13,22,25,26]

1. Center for Spatial and Functional Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
2. Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
3. Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
4. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
5. Department of Pathology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
6. Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, PA, USA
7. Scheie Eye Institute, Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, PA, USA
8. Penn Medicine Center for Ophthalmic Genetics in Complex Disease
9. Department of Orthopedic Surgery University of Michigan Medical School Ann Arbor, MI, USA
10. Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
11. Department of Child and Adolescent Psychiatry, Children's Hospital of Philadelphia, Philadelphia, PA, USA
12. Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
13. Institute for Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[14] Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[15] Department of Physiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[16] Department of Microbiology and Immunology, University of Melbourne, Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia

[17] Department of Critical Care, Melbourne Medical School, University of Melbourne, Melbourne, VIC, Australia

[18] Division of Infectious Diseases, Department of Medicine, University of British Columbia, Vancouver, BC, Canada

[19] Department of Surgery, University of British Columbia, Vancouver, BC, Canada

[20] BC Children's Hospital Research Institute, Vancouver, BC, Canada

[21] School of Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada

[22] Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[23] Department of Pathology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[24] Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[25] Division Endocrinology and Diabetes, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

[26] Penn Neurodegeneration Genomics Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Corresponding author: Struan F.A. Grant, email: grants@chop.edu

Word count: 4462

KEY POINTS

Question: What are the causal variants and corresponding effector genes conferring pediatric obesity susceptibility in different cellular contexts?

Findings: Our method of assessing 3D genomic data across a range of cell types revealed heritability enrichment of childhood obesity variants, particularly within pancreatic alpha cells. The mapping of putative causal variants to cis-regulatory elements revealed candidate effector genes for cell types spanning metabolic, neural, and immune systems.

Meaning: We gain a systemic view of childhood obesity genomics by leveraging 3D techniques that implicate regulatory regions harboring causal variants, providing insights into the disease pathogenesis across different cellular systems.

ABSTRACT

The prevalence of childhood obesity is increasing worldwide, along with the associated common comorbidities of type 2 diabetes and cardiovascular disease in later life. Motivated by evidence for a strong genetic component, our prior genome-wide association study (GWAS) efforts for childhood obesity revealed 19 independent signals for the trait; however, the mechanism of action of these loci remains to be elucidated. To molecularly characterize these childhood obesity loci we sought to determine the underlying causal variants and the corresponding effector genes within diverse cellular contexts. Integrating childhood obesity GWAS summary statistics with our existing 3D genomic datasets for 57 human cell types, consisting of high-resolution promoter-focused Capture-C/Hi-C, ATAC-seq, and RNA-seq, we applied stratified LD score regression and calculated the proportion of genome-wide SNP heritability attributable to cell type-specific features, revealing pancreatic alpha cell enrichment as the most statistically significant. Subsequent chromatin contact-based fine-mapping was carried out for genome-wide significant childhood obesity loci and their linkage disequilibrium proxies to implicate effector genes, yielded the most abundant number of candidate variants and target genes at the BDNF, ADCY3, TMEM18 and FTO loci in skeletal muscle myotubes and the pancreatic beta-cell line, EndoC-BH1. One novel implicated effector gene, ALKAL2 – an inflammation-responsive gene in nerve nociceptors – was observed at the key TMEM18 locus across multiple immune cell types. Interestingly, this observation was also supported through colocalization analysis using expression quantitative trait loci (eQTL) derived from the Genotype-Tissue Expression (GTEx) dataset, supporting an inflammatory and neurologic component to the pathogenesis of childhood obesity. Our comprehensive appraisal of 3D genomic datasets generated in a myriad of different cell types provides genomic insights into pediatric obesity pathogenesis.

INTRODUCTION

The prevalence of obesity has risen significantly worldwide[1], especially among children and adolescents[2]. Obesity is associated with chronic diseases, such as diabetes, cardiovascular diseases, and certain cancers[3-6], along with mechanical issues including osteoarthritis and sleep apnea[7].

Modern lifestyle factors, including physical inactivity, excessive caloric intake, and socioeconomic inequity, along with disrupted sleep and microbiome, represent environmental risk factors for obesity pathogenesis. However, genetics also play a significant role, with the estimated heritability ranging from 40% to 70%[8-10]. Studies show that body weight and obesity remain stable from infancy to adulthood[11-14], but variation between individuals does exisit[15]. Genome-wide association studies (GWAS) have improved our understanding of the genetic contribution to childhood obesity[16-21]. However, the functional consequences and molecular mechanisms of identified genetic variants in such GWAS efforts are yet to be fully elucidated. Efforts are now being made to predict target effector genes and explore potential drug targets using various computational and experimental approaches[22-26], which subsequently warrant functional follow-up efforts.

With our extensive datasets generated on a range of different cell types, by combining 3D chromatin maps (Hi-C, Capture-C) with matched transcriptome (RNA-seq) and chromatin accessibility data (ATAC-seq), we investigated heritability patterns of pediatric obesity-associated variants and their gene-regulatory functions in a cell type-specific manner. This approach yielded 94 candidate causal variants mapped to their putative effector gene(s) and corresponding cell type(s) setting. In addition, using methods comparable to our prior efforts in other disease contexts[27-34], we also uncovered new variant-to-gene combinations within specific novel cellular settings, most notably in immune cell types, which further confirmed the involvement of the immune system in the pathogenesis of obesity in the early stages of life.

METHODS

Data and resource: Datasets used in prior studies are listed in **eTable 1**. ATAC-seq, RNA-seq, Hi-C, and Capture-C *library generation* for each cell type is provided in their original published study and their *pre-processing pipelines and tools* can be found in **eMethods**.

Definition of cis-Regulatory Elements (cREs): We intersected ATAC-seq open chromatin regions (OCRs) of each cell type with chromatin conformation capture data determined by Hi-C/Capture-C of the same cell type, and with promoters (-1,500/+500bp of TSS) defined by GENCODE v30.

Childhood obesity GWAS summary statistics: Data on childhood obesity from the EGG consortium was downloaded from *www.egg-consortium.org*. We used 8,566,179 European ancestry variants (consisting of 8,613 cases and 12,696 controls in stage 1,921 cases and 1,930 controls in stage 2), representing ~55% of the total 15,504,218 variants observed across all ancestries in the original study[35]. The sumstats file was reformatted by *munge_sumstats.py* to standardize with the weighted variants from HapMap v3 within the LDSC baseline, which reduced the variants to 1,217,311 (7.8% of total).

Cell type specific partitioned heritability: We used *LDSC* (*http://www.github.com/bulik/ldsc*) *v.1.0.1* with *--h2* flag to estimate SNP-based heritability of childhood obesity within 4 defined sets of input genomic regions: (1) OCRs, (2) OCRs at gene promoters, (3) cREs, and (4) cREs with an expanded window of ±500 bp. The baseline model LD scores, plink filesets, allele frequencies and variants weights files for the European 1000 genomes project phase 3 in hg38 were downloaded from the provided link (*https://alkesgroup.broadinstitute.org/LDSCORE/GRCh38/*). The cREs of each cell type were

used to create the annotation, which in turn were used to compute annotation-specific LD scores for each cell types cREs set.

Genetic loci included in variant-to-genes mapping: 19 sentinel signals that achieved genome-wide significance in the trans-ancestral meta-analysis study[35] were leveraged for our analyses. Proxies for each sentinel SNP were queried using TopLD[36] and LDlinkR tool[37] with the GRCh38 Genome assembly, 1000 Genomes phase 3 v5 variant set, European population, and LD threshold of r^2>0.8, which resulted in 771 proxies, including the 21 SNPs from the 99% credible set of the original study (**eTable 2**).

GWAS-eQTL colocalization: The summary statistics for the European ancestry subset from the EGG consortium GWAS for childhood obesity was used. Common variants (MAF ≥ 0.01) from the 1000 Genomes Project v3 samples were used as a reference panel. We used non-overlapped genomic windows of ±250,000 bases extended in both directions from the median genomic position of each of 19 sentinel loci as input. We used ColocQuiaL[38] to test genome-wide colocalization of all possible variants included in each inputted window against GTEx v.8 eQTLs associations for all 49 tissues available from *https://www.gtexportal.org/home/datasets*. A conditional posterior probability of colocalization of 0.8 or greater was imposed.

RESULTS

Enrichment assessment of childhood obesity variants across cell types

To explore the enrichment of childhood obesity GWAS variants across cell types, we carried out Partitioned Linkage Disequilibrium Score Regression (LDSR)[39] on all ATAC-seq-defined OCRs for each cell type (**Fig. 1A)**. We observed that 41 of 57 cell types – including 22 metabolic, 21 immune, 7 neural cell types and 7 independent cell lines (**eTable 1**) – showed at least a degree of directional enrichment with the total set of OCRs (**Fig. 1B.a**). However, only four cell types – two pancreatic alpha and two pancreatic beta cell-based datasets – had statistically significant enrichments ($P<0.05$). These enrichments were less pronounced when focusing on promoter OCRs only (**Fig. 1B.b**). To further limit the LD enrichment assessment to just those OCRs that can putatively regulate gene expression via chromatin contacts with gene promoters, we used the putative cREs[27,29]. This reduced the number of cell types showing at least nominal enrichment (31 of 57), enlarged the dispersion of enrichment ranges across different cell types, increased the 95% confidence intervals (CI) of enrichments, and hence increased the *P*-value of the resulting regression score. cREs from pancreatic alpha cells derived from single-cell ATAC-seq were the only dataset that remained statically significant (**Fig. 1B.c**).

The original reported LDSR method analyzed enrichment in the 500bp flanking regions of their regulatory categories[39]. However, when we expanded our analysis to the ±500bp window for our cREs, albeit incorporating more weighted variants into the enrichment (represented by larger dots in **Fig. 1B.d**), this resulted in a decrease in the number of cell types yielding at least nominal enrichment (26 cell types), the enrichment range across cell types, the 95% CI, and level of significance. The pancreatic alpha cell observation also dropped below the bar for significance with this expanded window definition.

Consistency and diversity of childhood obesity proxy variants mapped to cREs

Despite the enrichments above only being limited to just a small number of cell types, it is likely that individual loci have differing levels of contributions in various cellular contexts and could not be detected at the genome wide assessment scale. As such we elected to further explore the candidate effector genes that are directly affected by cREs harboring childhood obesity-associated variants by systematically mapping the genomic positions of the LD proxies onto each cell type's cREs, resulting in several scenarios, as illustrated in **Fig. 2A,B**. The resulting 94 proxies clustered at 13 original loci (**eTable 3**). **eFig. 1A** outlines the number of signals at each locus included or excluded based on the criteria we defined for our regions of interest. The *TMEM18* locus yielded the most variants through cREs mapping, with 46 proxies for the two lead independent variants, rs7579427 and rs62104180. The second most abundant locus was *ADCY3*, with 21 proxies for lead variant rs4077678 (**Fig. 2C**). The higher number of variants at one locus did not correlate with implicating more genes or cell types through mapping. The mapping frequency of various variants within a specific locus exhibited substantial differences.

Inspecting individual variants regardless of their locus, we found that 28 of 94 proxies appeared in cREs across multiple cell types, with another 66 observed in just one cell type (**Fig. 2D)**. 45 variants of these 66 just contacted one gene promoter, such as at the *GPR1* and *TFAP2B* loci (**eFig. 2**).

Overall, the number of cell types in which a variant was observed in open chromatin correlated with the number of genes contacted via chromatin loops (**eFig. 3A**). However, we also observed that some variants found in cREs in multiple cell types were more selective with respect to their candidate effector genes (**eFig. 3B-red arrow**), or conversely, more selective across given cell types but implicated multiple genes (**eFig. 3B-blue arrow**). **eFig. 4** outlines our observations at the *TMEM18* locus – an example locus involved in both scenarios.

Implicated genes cluster at loci strongly associated with childhood obesity consistently across multiple cell types

Mapping the variants across all the cell types resulted in a total of 111 implicated childhood obesity candidate effector genes (**Table 1**). Among these, 45 genes were specific to just one cell type (**eFig. 5A**), including 13 in myotubes and 7 in natural killer cells. Conversely and notably, *BDNF* appeared across 42 different cell types. Across the metabolic, neural, and immune systems and seven other cell lines, there were 9 genes consistently implicated in all four categories (top panel **Fig. 3 –** red stars, **eFig. 5B**: "all"), while 5 genes were consistently implicated in metabolic, neural, and immune systems (top panel **Fig. 3** – blue stars, **eFig. 5B**: "all_main"). Two genes, *ADCY3* and *BDNF*, had variants both at their promoters and contacted variants in cREs via chromatin loops (**eFig. 6**).

At the *TMEM18* locus on chr 2p25.3, a highly significant human obesity locus that has long been associated with both adult and childhood obesity, we obeserved differing degrees of evidence for 16 genes, but noted that rs6548240, rs35796073, and rs35142762 consistently contacted the *SH3YL1, ACP1*, and *ALKAL2* promoters across multiple cell types (**Fig. 2D-third and fourth column**).

At the chr 2p23 locus, *ADCY3* yielded the most contacts (i.e. many proxies contacting the same gene via chromatin loops), suggesting this locus acts as a regulatory hub. However, we observed a similar composition in cell types for four other genes: *DNAJC27*, *DNAJC27-AS1* (both previously implicated in obesity and/or diabetes traits[40]), *AC013267.1*, and *SNORD14* (*RF00016*). *ITSN2*, *NCOA1*, and *EFR3B* were three genes within this locus that were only implicated in immune cell types. *NCOA1* encodes a prominent meta-inflammation factor[41] known to reduce adipogenesis and shift the energy balance between white and brown fat, and its absence known to induce obesity[42].

CALCR was the most frequently implicated gene at its locus, supported by 20 cell types across all systems. While within the *BDNF* locus, *METTL15* and *KIF18A* – two non-cell-type-

specific genes - plus some lncRNA genes, were contacted by childhood obesity-associated proxies within the same multiple cell types as *BNDF*, again suggesting the presence of a regulatory hub.

At the *FAIM2* locus on chr 12q13.12, we observed known genes associated with obesity, eating patterns, and diabetes-related traits, including *ASIC1*, *AQP2*, *AQP5*, *AQP6, RACGAP1*, and *AC025154.2 (AQP5-AS1)* along with *FAIM2* **(Table 1)**. These genes were harbored within PIR-OCRs of astrocytes, neural progenitors, hypothalamic neurons, and multiple metabolic cell types. Plasmacytoid and CD1c+ conventional dendritic cells were the only two immune cell types that harbored such proxies within their cREs, implicating *ASIC1*, *PRPF40B*, *RPL35AP28*, *TMBIM6*, *and LSM6P2* at the *FAIM2* locus.

The independent *ADCY9* and *FTO* loci are both located on chromosome 16. Genes at the *ADCY9* locus were only implicated in a subset of immune cell types. Interestingly, genes at the *FTO* locus were only implicated in Hi-C datasets (as opposed to Capture C), including 6 metabolic cell types and astrocytes. Most genes at the *FTO* locus were implicated in skeletal myotubes, differentiated osteoblasts, and astrocytes, namely *FTO* and *IRX3*; while *IRX5, CRNDE*, and *AC106738.1* were also implicated in adipocytes and hepatocytes.

The most implicated cell types by two sets of analyses

EndoC-BH1 and myotubes are the two cell types in which we implicated the most effector genes, with 38 and 42, respectively – **Fig. 3** side panel. This phenomenon is likely proportional in the case of myotubes, given the large number of cREs identified by overlapped Hi-C contact data and ATAC-seq open regions (**Fig. 1A**), but not for EndoC-BH1. Albeit harboring an average number of cREs compared to other cell types, EndoC-BH1 cells were consistently among the top-ranked heritability estimates for the childhood obesity variants resulting from the EGG consortium GWAS (**Fig. 1**) and harbored a significant number of implicated genes by the mapping of proxies. Interestingly, the pancreatic alpha cell type – shown above to be the most

significant for heritability estimate by LDSC – revealed only 6 implicated genes contacted by the defined proxies, namely *BDNF* and five lncRNA genes.

Pathway analysis

Of the 111 implicated genes in total, PubMed query revealed functional studies for 66 genes. The remaining were principally lncRNA and miRNA genes with currently undefined functions (**Table 1**). To investigate how our implicated genes could confer obesity risk, we performed several pathway analyses keeping them either separated for each cell type or pooling into the respective metabolic, neural, or immune system sets. **eFig. 7** shows simple Gene Ontology (GO) biological process terms enrichment results.

Leveraging the availability of our expression data generated via RNA-seq (available for 46 of 57 cell types), we performed pathway analysis. Given that our gene sets from the variant-to-gene process was stringently mapped, the sparse enrichment from normal direct analyses is not ideal for exploring obesity genetic etiology. Thus, we incorporated two methods from the *pathfindR* package[43] and our customized *SPIA* (details in **eMethods**). The result of 60 enriched KEGG terms is shown in **eFig. 8** (**eTable 4**), with 13 genes in 14 cell types for *pathfindR* and 39 enriched KEGG terms shown in **eFig. 9** (**eTable 5**), with 10 genes in 42 cell types for customized *SPIA*. There were 20 overlapping pathways between the two approaches (yellow rows in **eTable 4&5**) including many signaling pathways such as the GnRH (hsa04912), cAMP (hsa04024), HIF-1 (hsa04066), Glucagon (hsa04922), Relaxin (hsa04926), Apelin (hsa04371), and Phospholipase D (hsa04072) signaling pathways. They were all driven by one or more of these 5 genes: *ADCY3*, *ADCY9*, *CREBBP*, *MMP2,* and *NCOA1*. Interestingly, we observed the involvement of natural killer cells in nearly all the enriched KEGG terms from *pathfindR* due to the high expression of the two adenylyl cyclase encoded genes, *ADCY3* and *ADCY9*, along with *CREBBP*. The *SPIA* approach disregarded the aquaporin genes (given they appear so frequently in so many pathways that involve cellular channels) but highlighted the central role of

BDNF which single-handedly drove four signaling pathways: the Ras, Neurotrophin, PI3K-Akt, and MAPK signaling pathways. This also revealed the role of *TRAP1* in neurodegeneration.

These two approaches did not discount the role of *FAIM2* and *CALCR*. However, their absence was mainly due to the content of the current KEGG database. On the other hand, these approaches accentuated the role of the *MMP2* gene at the *FTO* locus in skeletal myotubes, given its consistency within the GnRH signaling pathway (**eFig. 10**), which is in line with previous studies linking its expression with obesity[44-46].

Supportive evidence by colocalization of target effector genes with eQTLs

The GTEx consortium has characterized thousands of eQTLs, albeit in heterogeneous bulk tissues[47]. To assess how many observed gene-SNP pairs agreed with our physical variant-to-gene mapping approach in our multiple separate cellular settings, we performed colocalization analysis using ColocQuiaL[38].

282 genes were reported to be associated with the variants within 13 loci from our variant-to-genes analysis. We found 114 colocalizations for ten of our loci that had high conditioned posterior probabilities (cond.PP.H4.abf ≥ 0.8), involving 44 genes and 41 tissues among the eQTLs. We extracted the posterior probabilities for each SNP within each colocalization and selected the 95% credible set as the likely causal variants (complete list in **eTable 6**). Despite sensitivity differences and varying cellular settings, when compared with our variant-to-gene mapping results, colocalization analysis yielded consistent identification for 21 pairs of SNP-gene interactions when considering the analyses across all our cell types, composed of 20 SNPs and 7 genes. Details of these SNP-gene pairs are shown in **Fig. 4A** and **B**.

Of these 20 SNPs, 15 were at the *ADCY3* locus, in LD with sentinel variant rs4077678, and all implicated *ADCY3* as the effector gene in 29 cell types – 15 metabolic, 6 immune, 4 neural cell types and 4 independent cell lines (**Fig. 4C**). Indeed, missense mutations have been

previously reported for this gene in the context of obesity[48,49] while another member of this gene family, *ADCY5*, has also been extensively implicated in metabolic traits[50].

Predicting transcription factors (TFs) binding disruption at implicated genes contributing to obesity risk

TFs regulate gene expression by binding to DNA motifs at enhancers and silencers, where any disruption by a SNP can potentially cause dysregulation of a target gene. Thus, we used *motifbreakR* (R package) to predict such possible events at the loci identified by our variant-to-gene mapping. Each variant was predicted to disrupt the binding of several different TFs, thus requiring further literature cross-examination to select the most probable effects. For example, rs7132908 (consistently contacting *FAIM2* in 25 cell types) was predicted to disrupt the binding of 12 different transcription factors. Among them, SREBF1 (**eFig. 11A**) was the only TF that concurred with evidence that it regulates *AQP2* and *FAIM2* at the same enhancer[51]. The full prediction list can be found in **eTable 7**.

To narrow down the list of putative TF binding sites at each variant position, we leveraged the ATAC-seq footprint analysis using the RGT suite[52]. The final set of Motif-Predicted Binding Sites (MPBS) within each cell type ATAC-seq footprints was used to overlap with the genomic locations of the OCRs, and then overlapped with our obesity variants, resulting in annotated 29 variants. Mosaic plot in **eFig. 11B** shows the number and proportions of variants predicted by *motifbreakR* and/or overlapped with MPBS. Insignificant *P*-value from Fisher's exact test indicated the independence of the two analyses. Only seven variants were found within the cREs for the same TF motifs predicted to be disrupted by *motifbreakR* (**eFig. 11C**). **eFig. 12** outlines the seven variants that *motifbreakR* and ATAC-seq footprint analysis agreed on the TF bindings they might disrupt.

DISCUSSION

Given the challenge of uncovering the underlying molecular mechanisms driving such a multifactorial disease as obesity, our approach leveraging GWAS summary statistics, RNA-seq, ATAC-seq, and promoter Capture C / Hi-C offers new insights. This is particularly true as it is becoming increasingly evident that multiple effector genes can operate in a temporal fashion at a given locus depending on cell state, including at the *FTO* locus[53]. Our approach offers an opportunity to implicate relevant cis-regulatory regions across different cell types contributing to the genetic etiology of the disease. By assigning GWAS signals to candidate causal variants and corresponding putative effector genes via open chromatin and chromatin contact information, we enhanced the fine-mapping process with an experimental genomic perspective to yield new insights into the biological pathways influencing childhood obesity.

LD score regression is a valuable method that estimates the relationship between linkage disequilibrium score and the summary statistics of GWAS SNPs to quantify the separate contributions of polygenic effects and various confounding factors that produce SNP-based heritability of disease. The general positive heritability enrichment across our open chromatin features spanning multiple cell types (**Fig. 1B.a**) reinforces the notion that obesity etiology involves many systems in our body.

While obesity has long been known to be a risk factor for pancreatitis and pancreatic cancer, the significant enrichment of pancreatic alpha and beta cell related 3D genomic features for childhood obesity GWAS signals demonstrates the bidirectional relationship between obesity and the pancreas; indeed, it is well established that insulin has obesogenic properties. Moreover, the comorbidity of obesity and diabetes (either causal or a result of the overlap between SNPs associated with these two diseases) is tangible. When focusing on genetic annotation of the cREs only, the association with obesity became more diverse across cell types, especially in metabolic cells. Interestingly, the lack of enrichment (only 8 of 57 cell types yielded no degree of enrichment) of obesity SNPs heritability in open gene promoters (**Fig.**

1B.b) reveals that cRE regions harboring obesity SNPs are more involved in gene regulation than disruption, and therefore potentially contributing more weight to the manifestation of the disease.

Of course, we should factor in the effective sample sizes of the GWAS efforts that are wide-ranging (2,000-24,000 – given that the N for each variant is different within a single dataset, thus contributing to the weights and *P*-value of each SNP when the algorithm calculates the genome-wide heritability), which could result in noise and negative enrichment observed in the analysis – a methodology limitation of partial linkage regression that has been extensively discussed in the field[54]. Thus, it is crucial to interpret the enrichment (or lack thereof) of disease variants in a certain cellular setting with an *ad hoc* biological context.

From mapping the common proxies of 19 independent sentinel SNPs that were genome-wide significantly associated with childhood obesity to putative effector genes through chromatin contacting cREs, one striking finding was the several potential "hubs" of putatively core effector genes, whose occurrence spread across three human physiological systems. With the data available from so many cell types, our approach connected new candidate causal variants to known obesity-related genes and new implications of cell modality for previously known associations.

A potential application of this association could be to fine-tune the effect of a drug toward controlling appetite. An example of bringing new aspects to the old is for the signal within the *FTO* locus that contacted *IRX3* and *IRX5*: previous studies have suggested these obesogenic effects operate in adipocytes[55], brain[56], or pancreas[57]; here we confirmed this association in adipocytes and uncover the presence of distal chromatin contacts in myotubes for the first time.

Besides the above-mentioned genes with known associations with obesity, we discovered newly implicated genes. For example, the *LRRIQ3* gene at the *TNNI3K* locus had its open promoter contacted by two SNPs, rs1040070 and rs10493544, in NTERA2 cells only. The published studies[58,59] that associated *LRRIQ3* with major depressive disorder and opioid usage

acknowledged the overlapping promoter of this gene, albeit in the opposite direction, with a run-through transcript of *FPGT-TNNI3K* – previously shown to be associated with BMI in European[60] and Korean populations[61].

It is apparent that not all the implicated genes we report would contribute equally to the susceptibility of obesity pathogenesis. Each locus comprises genes whose functions are obviously related to obesity or similar traits like BMI, fat weight, etc., while other genes are not so directly obvious in their relation to these traits.

It is encouraging that for implicated genes within these multi-cell-type loci across different physiological systems we could find previous associations to the corresponding cell types or systems. Examples are the two aforementioned genes at the *TMEM18* locus (*SH3YL1* and *ACP1*)[62-66] with the broad spectrum of their functions, *HEPACAM2* implicated in the NCIH716 cell line at the *CALCR* locus[67,68], and *LRRIQ3* in the NTERA2 cell line at the*TNNI3K* locus[69].

Chronic inflammation is an essential characteristic of obesity pathogenesis. Adipose tissue-resident immune cells have been observed, leading to an increased focus in recent years on their potential contribution to metabolic dysfunction. On the other hand, neurological or psychological conditions, such as stress, induce the secretion of both glucocorticoids (increase motivation for food) and insulin (promotes food intake and obesity). Pleasure feeding then reduces activity in the stress-response network, reinforcing the feeding habit. It has been shown that voluntary behaviors, stimulated by external or internal stressors or pleasurable feelings, memories, and habits, can override the basic homeostatic controls of energy balance[70]. The potential link between the immune system and metabolic disease, and moreover, through the neural system, was tangible in our findings.

Two of the three SNPs which ranked the third most consistent in our variant-to-gene mapping (**Fig. 2C**) – rs35796073 and rs35142762 – contacted the *ALKAL2* promoter (supported by GTEx evidence to colocalize with *ALKAL2* expression). The anaplastic lymphoma kinase (encoded by *ALK* gene) is a receptor tyrosine kinase, belongs to the insulin receptor family, and

has been reported to promote nerve cell growth and differentiation[71,72]. Despite *ALKAL2* (ALK and LTK ligand 2) being studied principally in the context of immunity, a recent study using the EGCUT biobank GWAS identified *ALK* as a candidate thinness gene and genetic deletion showed that its expression in hypothalamic neurons acts as a negative regulator in controlling energy expenditure via sympathetic control of adipose tissue lipolysis[73]. *ALKAL2* – encoding a high-affinity agonist of *ALK*/*LTK* receptors – which has been reported to enhance expression in response to inflammatory pain in nociceptors[74,75] - has been recently implicated as a novel candidate gene for childhood BMI by transcriptome-wide association study[76], and achieved genome-wide significance in a GWAS study contrasting persistent healthy thinness with severe early-onset obesity using the STILTS and SCOOP cohorts[77]. The finding that overexpression of *ALKAL2* could potentiate neuroblastoma progression in the absence of ALK mutation[78] echoes the relationship between *ADCY3* and *MC4R*[79], where a peripheral gene, *ADCY3*, can regulate/impair the function of a core gene, i.e. *MC4R*, within the energy-regulating melanocortin signaling pathway[80].

Our approach implicates putative target genes based on a mechanism of regulation for these variants to alter gene expression – through regulator TF(s) that bind to these contact sites. A potential limitation of the predictions from *motifbreakR* and matching TF motifs to ATAC-seq footprint by the RGT toolkit is that they were both based on the position probability matrixes of Jaspar and Hocomoco, which come from public motif databases. The ATAC-seq footprint analysis also carries sequence bias that can lead to false positive discovery. Thus, our attempt to call such regulators by predicting TF binding disruption can only serve as nominations – but warrant further functional follow up.

Another limitation of this work is the diversity in data quality among different samples, since different datasets were sampled and collected at different time points, from different patients, using different protocols, with libraries sequenced at different depths and qualities, and initially preprocessed with different pipelines and parameters. Thus, it is crucial to keep in mind that the

discrepancy in data points might have resulted from variations in data quality. Importantly, any association discovered must be validated functionally before effector genes of the genetic variants can be leveraged to develop new therapies. Their putative function(s) must be characterized, together with the mechanism whereby the given variant's alleles differentially affect the expression of the targeted genes. The next step is to explore how the target genes affect the trait of interest more directly.

Our results have provided a set of leads for future exploratory experiments in specific cellular settings in order to further expand our knowledge of childhood obesity genomics and hence equip us with more effective means to overcome the burden of this systematic disease.

CONCLUSION

Our approach of combining RNA-seq, ATAC-seq, and promoter Capture C/Hi-C datasets with GWAS summary statistics offers a systemic view of the multi-cellular nature of childhood obesity, shedding light on potential regulatory regions and effector genes. By leveraging physical properties, such as open chromatin status and chromatin contacts, we enhanced the fine-mapping process and gained new insights into the biological pathways influencing the disease. Although further functional validation is required, our findings provide valuable leads together with their cellular contexts for future research and the development of more effective strategies to address the burden of childhood obesity.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health awards R01 HD056465, R01 DK122586 and UM1 DK126194, and the Daniel B. Burke Endowed Chair for Diabetes Research.

Given the use of de-identified datasets and biospecimens was not considered human subjects research, ethical oversight was waived by the Institutional Review Board of the Children's Hospital of Philadelphia

TABLE AND FIGURE LEGENDS

Figure 1: Partitioned Linkage Disequilibrium Score Regression analysis for open chromatin regions of all cell types.

A. Bar-plot shows total number of open chromatin regions (OCRs) identified by ATAC-seq for each cell type on bulk cells - blue, or on single cell – red; and portion of OCRs that fall within promoter-interacting region (PIR-OCRs) identified by hi-C – green, or by capture-C – orange.

B. 4 panels of dot-plots show heritability enrichment by LDSC analysis for each cell type, with standard error whiskers. Dots' colors corresponding to p-values, dots with white asterisk are significant p-value<0.05, dots' sizes corresponding to proportion of SNP contribute for heritability. Dash line at 1, i.e., no enrichment.

a) Analysis done on whole OCRs set of each cell type (whiskers colors match with bulk/single cell from bar-plot A);
b) On only OCRs that overlapped with promoters (whiskers' colors match with bulk/single cell from bar-plot A);
c) On the putative cREs of each cell type (whiskers' colors match with hi-C/capture-C from bar-plot A);
d) On the same PIR-OCRs as middle panel with their genomic positions expanded ±500 bases on both sides (whiskers' colors match with hi-C/capture-C from bar-plot A).

Figure 2: Mapping 771 proxies to the open chromatin regions of each cell type

A. Schematic shows 4 different ways OCRs can be subset due to overlapping with chromatin contact sites and gene promoters, colors of sets of OCRs applied for the areas in B panel.
B. Venn diagram shows how 771 proxies mapped onto the OCRs:
   - Blue area: 758 proxies were located within contact regions of at least one cell type regardless of chromatin state;
   - Red area: 417 proxies were located within contact regions marked as open by overlapping with OCR;
   - Yellow area: If we only considered open chromatin regions, 178 proxies were included;

- Green area: To focus on just those variants residing within open chromatin and contacting promoter regions in any cell type, we overlapped the genomic positions of these proxies with each cell type's cRE set, yielding 90 variants (3 from the 99% credible set) directly contacting open gene promoters **(eTable 3)**, with 10 located within a promoter of one gene but contacting another different gene promoter. There were an additional 4 variants located within gene promoters but in chromatin contact with promoter(s) of nearby transcript(s) of the same gene.

C. Bar-plot shows number of proxies, cell types and target genes mapped at each locus.

D. The upSet plot shows the degree of overlap across cell types of the variants; ranked from the most common variant (red) – rs61888800 from *BDNF* locus, a well-known 5' untranslated region variant of this gene that is associated with anti-depression and therapeutic response[81,82] – appeared in 39 cell types, to the group of variants (grey) which appeared in only one cell type.

Figure 3: Profiles of 111 implicated genes by 94 proxies through cREs of each cell type

Main panel: Bubble plot show corresponding expression level (size) and number of variants (color) target each implicated gene of each cell type. Squares represent genes with variants at their promoters. Circles represent genes with variants contacted through chromatin loops. Some genes were implicated by both types, these "double implications" are represented as diamond shapes, and were identified across several cell types: two cell types (plasmacytoid dendritic cells and pre-differentiated adipocytes) for *ADCY3* gene, and five for *BDNF* (human embryonic stem cells - hESC, differentiated human fetal osteoblast cells - hFOB_Diff, neural progenitor cells derived from induced pluripotent stem cells - NPC_iPSC, PANC-1, and NCIH716 cell lines) Genes with expression undetected in our arrays are shown as triangles.

Top panel: bar-plot shows numbers of cell types each gene was implicated within, color-coded by which systems the cell types belong to.

Right panel: bar-plot shows numbers of genes implicated by the variants with each cell type.

Figure 4: Colocalization of target effector genes with eQTLs

A. Venn diagram shows the overlaps between sets of genes yielded by ColocQuiaL and the variant-to-gene mapping process.

B. Circos plot of the 10 loci demonstrates the differences in the ranges of associations between the two approaches, with long-ranged chromatin contacts between obesity variants and target genes displayed as orange links and short-range eQTLs colocalizations as green links.

Two SNPs – rs35796073, and rs35142762 within the *TMEM18* locus, in linkage disequilibrium with rs7579427 – were estimated with high probability (cond.PP.H4=0.78) of colocalizing with the expression of *ALKAL2* gene in subcutaneous adipose tissue. These pairs of SNP-gene were also identified by our variant-to-gene mapping approach in natural killer cells, plasmacytoid dendritic cells, unstimulated PBMC naïve CD4 T cells and astrocytes.

The rs7132908 variant at the *FAIM2* locus colocalized with the expression of *AQP6* in thyroid tissue and with *ASIC1* in prostate tissue, not only with high cond.PP.H4 but also with high individual SNP causal probability (SNP.PP.H4 > 0.95). rs7132908 was the second most consistent observation in our variant-to-gene mapping, namely across 25 different cell types (**Figure 2C**) and all three systems plus the other independent cell lines. The pair of rs7132908-contacting-*AQP6* was observed in 15 different cell types - 8 metabolic and 4 neural cell types, and 3 independent cell lines. The pair of rs7132908-contacting-*ASIC1* was observed in 11 different cell types - 8 metabolic and 2 neural cell types, and plasmacytoid dendritic cells.

The other eQTL signals that overlapped with our variant-to-gene mapping results were: *BDNF* at the *METTL15* locus with its promoter physically contacted by rs11030197 in 4 cell types and its expression significantly colocalized (cond.PP.H4=0.82) in tibial artery; *ADCY9* at its locus with its promoter physically contacted by rs2531995 in natural killer cells and its expression significantly colocalized in skin tissue ("Skin_Not_Sun_Exposed_Suprapubic", cond.PP.H4=0.97). And *ADCY3* in the C panel.

C. ColocQuiaL estimated that these SNPs highly colocalize with the expression of *ADCY3* in 11 different tissues, where the overlapping with the 16 cell types is represented**,** color-coded by the proxies rs numbers.

Table 1 – PubMed-query known functions for 111 genes implicated by obesity variants.

Locus	Implicated genes	Obesity or related traits	Different traits

TNNI3K	LRRIQ3	(NA)	Associated with opioid usage [PMID:34728798] and MDD [PMID: 31748543]
	FPGT	Predict BMI in Korean pop.[PMID: 28674662]	(NA)
	FPGT-TNNI3K		Associated with MDD [PMID: 31748543]
	LRRC53	Associated with high BMI increased risk heart attack [PMID: 32471361]	(NA)
	ASTN1	Identified as obesity QTL in rat [PMID: 35729251]	Associated with neurodevelopmental traits [PMID: 24381304] and variety of cancers [PMID: 32945491]
	BRINP2	(NA)	Associated with neurodevelopmental traits [PMID: 34267256]
SEC16B	AL122019.1	(NA)	
	AL162431.1		
TMEM18	FAM110C	(NA)	Overexpression induces microtubule aberrancies [PMID: 17499476], involved in cell spreading and migration [PMID: 19698782]
	SH3YL1	Associated with BMI in type 2 diabetes nephropathy [PMID: 33223406]	Influence on T cell activation [PMID: 31427643], involved in different cancer types [PMID: 26305679,24508479]
	ACP1	Associated with early-onset obesity [PMID: 24129437], correlated with cardiovascular risks [PMID: 19570551], drive adipocyte differentiation via control of pdgfrα signaling [PMID: 33615467]	Associated with bipolar disorder [PMID: 31830721]
	ALKAL2	Associated with childhood BMI [PMID: 33627773]	Enhance expression in response to inflammatory pain in nociceptors [PMID: 35608912, 35610945]
	MYT1L	Associated with early-onset obesity [PMID: 24129437]	(NA)
	AC079779.1	(NA)	
	AC079779.2		
	AC079779.3		
	AC079779.4		
	LINC01865		
	AC105393.2		
	AC105393.1		

	LINC01874		
	LINC01875		
	AC093326.1		
	AC141930.2		
ADCY3	ITSN2	(NA)	Regulate T-cells function [PMID: 32618424] and help the interaction with B-cells [PMID: 29337666]
	NCOA1	Meta-inflammation gene [PMID: 25647480], reduce adipogenesis, shift the energy balance between white and brown fat [PMID: 31133421]	(NA)
	ADCY3	Regulate/impair MC4R within energy-regulating melanocortin signaling pathway [PMID: 29311635,32955435]	
	DNAJC27-AS1	Linked to obesity, diabetes traits [PMID: 30131766]	
	DNAJC27	Linked to obesity, diabetes traits [PMID: 30131766]	
	EFR3B	Associated with T1D [PMID: 21980299], down-regulated in rare obesity-related disorder [PMID: 25705109]	
	WDR43	(NA)	Associated with breast cancer [PMID: 27117709]
	AC013267.1	(NA)	
	RF00016		
GPR1	GPR1	Increase expression in obese phenotype [PMID: 34174278]	(NA)
TFAP2B	TFAP2D	(NA)	Involve in embryogenesis [PMID: 12711551]
CALCR	HEPACAM2		Associated with colorectal cancer [PMID: 29659199, 29973580]
	VPS50		Involve in neurodevelopmental disorders and defects [PMID: 30828385, 34037727]
	MIR653		Involve in different types of cancer [PMID: 35777307]
	MIR489	Promote adipogenesis in mice [PMID: 34004251]	(NA)
	CALCR	Associated with BMI and control of food-intake [PMID: 34462445, 34210852, 31955990, 29522093]	
	TFPI2	(NA)	Involved in colorectal cancer [PMID: 35004840, 34092617, 25902909]

	BET1	Involved in triacylglycerol metabolism [PMID: 24423365]	Associated with muscular dystrophy [PMID: 34310943, 34779586]
	AC003092.1	(NA)	Association with glioblastoma [PMID: 33815468, 30442884]
	AC002076.1	(NA)	
BDNF	LIN7C	Associated in T2D [PMID: 20215397], obesity [PMID: 23044507]	Associated with psychopathology [PMID: 23044507]
	BDNF-AS	Regulate *BDNF* and *LIN7C* expression [PMID: 22960213, 22446693]	(NA)
	BDNF	Regulate eating behavior and energy balance [PMID: 34556834]	
	MIR610	(NA)	Involve in different types of cancer [PMID: 34408418, 29228616, 26885452]
	KIF18A		Involve in different types of cancer [PMID: 35591854, 35286090]
	METTL15	Associated with childhood obesity [PMID: 31504550]	(NA)
	AC090124.1	(NA)	Reported to differentially prognostic of pancreatic cancer [PMID: 34307375]
	ARL14EP		Involve in WAGR syndrome [PMID: 36011342, 31511512]
	DCDC1		Involvement with eyes anomalies [PMID: 34773354, 34703991]
	THEM7P		Associated with mechanisms underlying inguinal hernia [PMID: 34392144]
	AL035078.2	(NA)	
	ELP4		
	LINC00678		
	AC023206.1		
	RN7SKP158		
	AC104978.1		
	MIR8068		
	AC013714.1		
	AC100773.1		
	AC090833.1		
	AC090791.1		
	AC110056.1		
	AL035078.2		

FAIM2	PRPF40B	(NA)	Splicing regulator involved in T-cell development [PMID: 31088860, 34323272]
	TMBIM6	Deficiency leads to obesity by increasing Ca2+-dependent insulin secretion [PMID: 32394396]	Immune cell function and survival [PMID: 26470731]
	BCDIN3D	Associated with obesity, T2D [PMID: 20215397]	(NA)
	FAIM2	Associated with childhood obesity [PMID: 31504550]	
	AQP2	Associated with obesity, diabetes [PMID: 33367818]	
	AQP5	Associated with non-obese diabetes [PMID: 25635992,22320885]	Responsible for transporting water, involve in Sjogren's syndrome [PMID: 25635992, 31557796]
	AQP6	Down-regulated in retina in diabetes [PMID: 21851171]	Associated with renal diseases [PMID: 30654539]
	RACGAP1	Involve in diabetes nephropathy [PMID: 35222021]	(NA)
	ASIC1	Inhibition increase food intake and decrease energy expenditure [PMID: 35894166]	
	LSM6P2	(NA)	
	RPL35AP28		
	LINC02396		
	LINC02395		
	AC025154.1		
	AC025154.2		
ADCY9	SLX4	(NA)	Associated with blood pressure [PMID: 30671673]
	DNASE1	Associated with obesity hypertension [PMID: 33351325]	(NA)
	TRAP1	Involve in global metabolic network, deletion reduce obesity incidence [PMID: 25088416]	
	CREBBP	Associated with high adiposity and low cardiometabolic risk [PMID: 33619380]	
	ADCY9	Asoociated with BMI, obesity [PMID: 33619380, 23563607]	
	SRL	(NA)	Involve in cardiac dysfunction [PMID: 22119571]
	LINC01569		Associated with cancer and endometriosis [PMID: 35341703, 34422671]
	TFAP4	Associated with BMI, birth weight,	(NA)

		maternal glycemic [PMID: 35708509]	
	AC012676.1	(NA)	Involve in hepatocellular carcinoma [PMID: 35210216]
	AC009171.2	(NA)	
FTO	FTO	Most extensively studied obesity locus [PMID: 34556834]	(NA)
	IRX3	Obesogenic effects in adipocytes [PMID: 26760096], brain [PMID: 24646999], pancreas[93]	
	IRX5		
	AC018553.1	(NA)	Associated with melanoma [PMID: 35611195]
	CRNDE	Regulator of angiogenesis in obesity-induced diabetes [PMID: 31863035]	(NA)
	MMP2	Involve in obesity-relate angiogenesis [PMID: 35919566]	
	CAPNS2	(NA)	Associated with thyroid-related traits [PMID: 23408906]
	AMFR	Involve in hepatic lipid metabolism [PMID: 33591966]	(NA)
	CETP	Involve in monogenic hyperalphalipoproteinemia [PMID: 34878751]	
	RPGRIP1L	Hypomorphism of this ciliary gene linked to morbid obesity [PMID: 27064284, 30597647, 29657248]	Required for hypothalamic arcuate neuron development [PMID: 30728336]
	LINC02169	(NA)	Associated with occupational exposure to gases/fumes and mineral dust [PMID: 31152171]
	AC007491.1	(NA)	
	AC018553.2		
	LINC02140		
	AC106738.1		
	AC106738.2		
	MTND5P34		
	AC007336.1		
MC4R	AC090771.1	(NA)	

REFERENCES

1. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet*. Dec 16 2017;390(10113):2627-2642. doi:10.1016/s0140-6736(17)32129-3
2. National Health and Nutrition Examination Survey 2017–March 2020 Prepandemic Data Files Development of Files and Prevalence Estimates for Selected Health Outcomes, http://dx.doi.org/10.15620/cdc:106273 (2021). https://stacks.cdc.gov/view/cdc/106273
3. Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *New England Journal of Medicine*. 2017;377(1):13-27. doi:10.1056/NEJMoa1614362
4. Singh GM, Danaei G, Farzadfar F, et al. The Age-Specific Quantitative Effects of Metabolic Risk Factors on Cardiovascular Diseases and Diabetes: A Pooled Analysis. *PLOS ONE*. 2013;8(7):e65174. doi:10.1371/journal.pone.0065174
5. The Emerging Risk Factors C. Separate and combined associations of body-mass index and abdominal adiposity with cardiovascular disease: collaborative analysis of 58 prospective studies. *The Lancet*. 2011/03/26/ 2011;377(9771):1085-1095. doi:https://doi.org/10.1016/S0140-6736(11)60105-0
6. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer — Viewpoint of the IARC Working Group. *New England Journal of Medicine*. 2016;375(8):794-798. doi:10.1056/NEJMsr1606602
7. Fontaine KR, Barofsky I. Obesity and health-related quality of life. *Obes Rev*. Aug 2001;2(3):173-82. doi:10.1046/j.1467-789x.2001.00032.x
8. Loos RJF, Yeo GSH. The genetics of obesity: from discovery to biology. *Nature Reviews Genetics*. 2022/02/01 2022;23(2):120-133. doi:10.1038/s41576-021-00414-z
9. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behavior genetics*. Jul 1997;27(4):325-51.
10. Elks CE, Den Hoed M, Zhao JH, et al. Variability in the heritability of body mass index: a systematic review and meta-regression. *Frontiers in endocrinology*. 2012;3:29.
11. Demerath EW, Choh AC, Czerwinski SA, et al. Genetic and environmental influences on infant weight and weight change: the Fels Longitudinal Study. *Am J Hum Biol*. Sep-Oct 2007;19(5):692-702. doi:10.1002/ajhb.20660
12. Dubois L, Girard M, Girard A, Tremblay R, Boivin M, Pérusse D. Genetic and environmental influences on body size in early childhood: a twin birth-cohort study. *Twin Res Hum Genet*. Jun 2007;10(3):479-85. doi:10.1375/twin.10.3.479
13. Wardle J, Carnell S, Haworth CM, Plomin R. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. *Am J Clin Nutr*. Feb 2008;87(2):398-404. doi:10.1093/ajcn/87.2.398
14. Bouchard C. Childhood obesity: are genetic differences involved? *The American Journal of Clinical Nutrition*. 2009;89(5):1494S-1501S. doi:10.3945/ajcn.2009.27113C
15. Littleton SH, Berkowitz RI, Grant SFA. Genetic Determinants of Childhood Obesity. *Mol Diagn Ther*. Dec 2020;24(6):653-663. doi:10.1007/s40291-020-00496-1

16. Vogelezang S, Bradfield JP, Ahluwalia TS, et al. Novel loci for childhood body mass index and shared heritability with adult cardiometabolic traits. *PLOS Genetics*. 2020;16(10):e1008718. doi:10.1371/journal.pgen.1008718
17. Yaghootkar H, Zhang Y, Spracklen CN, et al. Genetic Studies of Leptin Concentrations Implicate Leptin in the Regulation of Early Adiposity. *Diabetes*. Dec 2020;69(12):2806-2818. doi:10.2337/db20-0070
18. Couto Alves A, De Silva NMG, Karhunen V, et al. GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI. *Sci Adv*. Sep 2019;5(9):eaaw3095. doi:10.1126/sciadv.aaw3095
19. Fu J, Wang Y, Li G, et al. Childhood sleep duration modifies the polygenic risk for obesity in youth through leptin pathway: the Beijing Child and Adolescent Metabolic Syndrome cohort study. *Int J Obes (Lond)*. Aug 2019;43(8):1556-1567. doi:10.1038/s41366-019-0405-1
20. Turcot V, Lu Y, Highland HM, et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet*. Jan 2018;50(1):26-41. doi:10.1038/s41588-017-0011-x
21. Littleton SH, Grant SFA. Strategies to identify causal common genetic variants and corresponding effector genes for paediatric obesity. *Pediatr Obes*. Dec 2022;17(12):e12968. doi:10.1111/ijpo.12968
22. Yu F, Cato LD, Weng C, et al. Variant to function mapping at single-cell resolution through network propagation. *Nature Biotechnology*. 2022/06/06 2022;doi:10.1038/s41587-022-01341-y
23. Avsec Ž, Agarwal V, Visentin D, et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nature Methods*. 2021/10/01 2021;18(10):1196-1203. doi:10.1038/s41592-021-01252-x
24. Gazal S, Weissbrod O, Hormozdiari F, et al. Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity. *Nature Genetics*. 2022/06/01 2022;54(6):827-836. doi:10.1038/s41588-022-01087-y
25. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning–based sequence model. *Nature Methods*. 2015/10/01 2015;12(10):931-934. doi:10.1038/nmeth.3547
26. Nasser J, Bergman DT, Fulco CP, et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature*. 2021/05/01 2021;593(7858):238-243. doi:10.1038/s41586-021-03446-x
27. Chesi A, Wagley Y, Johnson ME, et al. Genome-scale Capture C promoter interactions implicate effector genes at GWAS loci for bone mineral density. OriginalPaper. *Nature Communications*. 2019-03-19 2019;10(1):1-11. doi:doi:10.1038/s41467-019-09302-x
28. Su C, Johnson ME, Torres A, et al. Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. *Nature Communications*. 2020/07/03 2020;11(1):3294. doi:10.1038/s41467-020-17089-5
29. Pahl MC, Doege CA, Hodge KM, et al. Cis-regulatory architecture of human ESC-derived hypothalamic neuron differentiation aids in variant-to-gene mapping of relevant complex traits. OriginalPaper. *Nature Communications*. 2021-11-19 2021;12(1):1-12. doi:doi:10.1038/s41467-021-27001-4

30. Cousminer DL, Wagley Y, Pippin JA, et al. Genome-wide association study implicates novel loci and reveals candidate effector genes for longitudinal pediatric bone accrual. *Genome Biol*. Jan 4 2021;22(1):1. doi:10.1186/s13059-020-02207-9
31. Vujkovic M, Ramdas S, Lorenz KM, et al. A multiancestry genome-wide association study of unexplained chronic ALT elevation as a proxy for nonalcoholic fatty liver disease with histological and radiological validation. *Nat Genet*. Jun 2022;54(6):761-771. doi:10.1038/s41588-022-01078-z
32. Pahl MC, Le Coz C, Su C, et al. Implicating effector genes at COVID-19 GWAS loci using promoter-focused Capture-C in disease-relevant immune cell types. *Genome Biology*. 2022/06/03 2022;23(1):125. doi:10.1186/s13059-022-02691-1
33. Su C, Gao L, May CL, et al. 3D chromatin maps of the human pancreas reveal lineage-specific regulatory architecture of T2D risk. *Cell Metabolism*. 6 September 2022 2022;34(9)1409. doi:https://doi.org/10.1016/j.cmet.2022.08.014
34. Palermo J, Chesi A, Zimmerman A, et al. Variant-to-gene mapping followed by cross-species genetic screening identifies GPI-anchor biosynthesis as a regulator of sleep. *Sci Adv*. Jan 6 2023;9(1):eabq0844. doi:10.1126/sciadv.abq0844
35. Bradfield JP, Vogelezang S, Felix JF, et al. A trans-ancestral meta-analysis of genome-wide association studies reveals loci associated with childhood obesity. *Human Molecular Genetics*. 2019;28(19):3327-3338. doi:10.1093/hmg/ddz161
36. Huang L, Rosen JD, Sun Q, et al. TOP-LD: A tool to explore linkage disequilibrium with TOPMed whole-genome sequence data. *The American Journal of Human Genetics*. 2022/06/02/ 2022;109(6):1175-1181. doi:https://doi.org/10.1016/j.ajhg.2022.04.006
37. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Front Genet*. 2020;11:157. doi:10.3389/fgene.2020.00157
38. Chen BY, Bone WP, Lorenz K, Levin M, Ritchie MD, Voight BF. ColocQuiaL: a QTL-GWAS colocalization pipeline. *Bioinformatics*. 2022;38(18):4409-4411. doi:10.1093/bioinformatics/btac512
39. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. OriginalPaper. *Nature Genetics*. 2015-09-28 2015;47(11):1228-1235. doi:doi:10.1038/ng.3404
40. Cherian PT, Al-Khairi I, Sriraman D, et al. Increased Circulation and Adipose Tissue Levels of DNAJC27/RBJ in Obesity and Type 2-Diabetes. *Front Endocrinol (Lausanne)*. 2018;9:423. doi:10.3389/fendo.2018.00423
41. Rollins DA, Coppo M, Rogatsky I. Minireview: nuclear receptor coregulators of the p160 family: insights into inflammation and metabolism. *Mol Endocrinol*. Apr 2015;29(4):502-17. doi:10.1210/me.2015-1005
42. Mohsen G A-M, Abu-Taweel GM, Rajagopal R, et al. Betulinic acid lowers lipid accumulation in adipocytes through enhanced NCoA1–PPARγ interaction. *Journal of Infection and Public Health*. 2019/09/01/ 2019;12(5):726-732. doi:https://doi.org/10.1016/j.jiph.2019.05.011
43. Ulgen E, Ozisik O, Sezerman OU. pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks. Methods. *Frontiers in Genetics*. 2019-September-25 2019;10doi:10.3389/fgene.2019.00858

44. Derosa G, Ferrari I, D'Angelo A, et al. Matrix metalloproteinase-2 and -9 levels in obese patients. *Endothelium*. Jul-Aug 2008;15(4):219-24. doi:10.1080/10623320802228815
45. Sezgin SBA, Bayoglu B, Ersoz F, et al. Downregulation of MMP-2 and MMP-9 genes in obesity patients and their relation with obesity-related phenotypes. *Turkish Journal of Biochemistry*. 2022;47(4):425-433. doi:doi:10.1515/tjb-2021-0124
46. Nonino CB, Noronha NY, de Araújo Ferreira-Julio M, et al. Differential Expression of MMP2 and TIMP2 in Peripheral Blood Mononuclear Cells After Roux-en-Y Gastric Bypass. Brief Research Report. *Frontiers in Nutrition*. 2021-October-13 2021;8doi:10.3389/fnut.2021.628759
47. Mohammadi Pejman 5 6 Park YoSon 11 Parsana Princy 12 Segrè Ayellet V. 1 Strober Benjamin J. 9 Zappala Zachary 7 8 GCLaAFBAACSEDJRHYJB, P. 19 Volpi Simona 19 NpmAAGPKSLARLNCMHMRASJ, 16 PSLBMEBPA, 137 NCFNCR. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213.
48. Grarup N, Moltke I, Andersen MK, et al. Loss-of-function variants in ADCY3 increase risk of obesity and type 2 diabetes. *Nature genetics*. 2018;50(2):172-174. doi:10.1038/s41588-017-0022-7
49. Stergiakouli E, Gaillard R, Tavaré JM, et al. Genome-wide association study of height-adjusted BMI in childhood identifies functional variant in ADCY3. *Obesity (Silver Spring)*. Oct 2014;22(10):2252-9. doi:10.1002/oby.20840
50. Sinnott-Armstrong N, Sousa IS, Laber S, et al. A regulatory variant at 3q21.1 confers an increased pleiotropic risk for hyperglycemia and altered bone mineral density. *Cell Metabolism*. 2021/03/02/ 2021;33(3):615-628.e13. doi:https://doi.org/10.1016/j.cmet.2021.01.001
51. Kikuchi H, Jung HJ, Raghuram V, et al. Bayesian identification of candidate transcription factors for the regulation of Aqp2 gene expression. *Am J Physiol Renal Physiol*. Sep 1 2021;321(3):F389-f401. doi:10.1152/ajprenal.00204.2021
52. Li Z, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. Identification of transcription factor binding sites using ATAC-seq. *Genome Biology*. 2019/02/26 2019;20(1):45. doi:10.1186/s13059-019-1642-2
53. Sobreira DR, Joslin AC, Zhang Q, et al. Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of IRX3 and IRX5. *Science (New York, NY)*. 2021;372(6546):1085-1091. doi:10.1126/science.abf1008
54. Steinsaltz D, Dahl A, Wachter KW. On Negative Heritability and Negative Estimates of Heritability. *Genetics*. 2020;215(2):343-357. doi:10.1534/genetics.120.303161
55. Claussnitzer M, Hui CC, Kellis M. FTO Obesity Variant and Adipocyte Browning in Humans. *N Engl J Med*. Jan 14 2016;374(2):192-3. doi:10.1056/NEJMc1513316
56. Smemo S, Tena JJ, Kim K-H, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014/03/01 2014;507(7492):371-375. doi:10.1038/nature13138
57. Ragvin A, Moro E, Fredman D, et al. Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. *Proc Natl Acad Sci U S A*. Jan 12 2010;107(2):775-80. doi:10.1073/pnas.0911591107
58. Johnston KJA, Adams MJ, Nicholl BI, et al. Identification of novel common variants associated with chronic pain using conditional false discovery rate analysis with major depressive disorder and assessment of pleiotropic effects of LRFN5. *Transl Psychiatry*. Nov 20 2019;9(1):310. doi:10.1038/s41398-019-0613-4

59. Sanchez-Roige S, Fontanillas P, Jennings MV, et al. Genome-wide association study of problematic opioid prescription use in 132,113 23andMe research participants of European ancestry. *Mol Psychiatry*. Nov 2021;26(11):6209-6217. doi:10.1038/s41380-021-01335-3
60. Graff M, Ngwa JS, Workalemahu T, et al. Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course. *Hum Mol Genet*. Sep 1 2013;22(17):3597-607. doi:10.1093/hmg/ddt205
61. Lee JS, Cheong HS, Shin HD. BMI prediction within a Korean population. *PeerJ*. 2017;5:e3510. doi:10.7717/peerj.3510
62. Fernandes SJ, Morikawa H, Ewing E, et al. Non-parametric combination analysis of multiple data types enables detection of novel regulatory mechanisms in T cells of multiple sclerosis patients. *Sci Rep*. Aug 19 2019;9(1):11996. doi:10.1038/s41598-019-48493-7
63. Blessing AM, Ganesan S, Rajapakshe K, et al. Identification of a Novel Coregulator, SH3YL1, That Interacts With the Androgen Receptor N-Terminus. *Mol Endocrinol*. Oct 2015;29(10):1426-39. doi:10.1210/me.2015-1079
64. Kobayashi M, Harada K, Negishi M, Katoh H. Dock4 forms a complex with SH3YL1 and regulates cancer cell migration. *Cell Signal*. May 2014;26(5):1082-8. doi:10.1016/j.cellsig.2014.01.027
65. Choi GS, Min HS, Cha JJ, et al. SH3YL1 protein as a novel biomarker for diabetic nephropathy in type 2 diabetes mellitus. *Nutr Metab Cardiovasc Dis*. Feb 8 2021;31(2):498-505. doi:10.1016/j.numecd.2020.09.024
66. Gaynor SC, Monson ET, Gaine ME, et al. Male-specific association of the 2p25 region with suicide attempt in bipolar disorder. *J Psychiatr Res*. Feb 2020;121:151-158. doi:10.1016/j.jpsychires.2019.11.009
67. Wu Z, Liu Z, Ge W, et al. Analysis of potential genes and pathways associated with the colorectal normal mucosa-adenoma-carcinoma sequence. *Cancer Med*. Jun 2018;7(6):2555-2566. doi:10.1002/cam4.1484
68. Huang Z, Yang Q, Huang Z. Identification of Critical Genes and Five Prognostic Biomarkers Associated with Colorectal Cancer. *Med Sci Monit*. Jul 5 2018;24:4625-4633. doi:10.12659/msm.907224
69. Pleasure SJ, Lee VM. NTera 2 cells: a human cell line which displays characteristics expected of a human committed neuronal progenitor cell. *J Neurosci Res*. Aug 15 1993;35(6):585-602. doi:10.1002/jnr.490350603
70. Dallman MF. Stress-induced obesity and the emotional nervous system. *Trends Endocrinol Metab*. Mar 2010;21(3):159-65. doi:10.1016/j.tem.2009.10.004
71. Iwahara T, Fujimoto J, Wen D, et al. Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. *Oncogene*. Jan 30 1997;14(4):439-49. doi:10.1038/sj.onc.1200849
72. Motegi A, Fujimoto J, Kotani M, Sakuraba H, Yamamoto T. ALK receptor tyrosine kinase promotes cell growth and neurite outgrowth. *J Cell Sci*. Jul 1 2004;117(Pt 15):3319-29. doi:10.1242/jcs.01183
73. Orthofer M, Valsesia A, Mägi R, et al. Identification of ALK in Thinness. *Cell*. 2020/06/11/ 2020;181(6):1246-1262.e22. doi:https://doi.org/10.1016/j.cell.2020.04.034

74. Defaye M, Iftinca MC, Gadotti VM, et al. The neuronal tyrosine kinase receptor ligand ALKAL2 mediates persistent pain. *The Journal of Clinical Investigation*. 06/15/ 2022;132(12)doi:10.1172/JCI154317
75. Sun W, Yang S, Wu S, et al. Transcriptome analysis reveals dysregulation of inflammatory and neuronal function in dorsal root ganglion of paclitaxel-induced peripheral neuropathy rats. *Mol Pain*. May 24 2022:17448069221106167. doi:10.1177/17448069221106167
76. Yao S, Wu H, Ding J-M, et al. Transcriptome-wide association study identifies multiple genes associated with childhood body mass index. *International Journal of Obesity*. 2021/05/01 2021;45(5):1105-1113. doi:10.1038/s41366-021-00780-y
77. Riveros-McKay FA-O, Mistry V, Bounds RA-O, et al. Genetic architecture of human thinness compared to severe obesity. 2019;(1553-7404 (Electronic))
78. Borenäs M, Umapathy G, Lai W-Y, et al. ALK ligand ALKAL2 potentiates MYCN-driven neuroblastoma in the absence of ALK mutation. *The EMBO Journal*. 2021;40(3):e105784. doi:https://doi.org/10.15252/embj.2020105784
79. Siljee JE, Wang Y, Bernard AA, et al. Subcellular localization of MC4R with ADCY3 at neuronal primary cilia underlies a common pathway for genetic predisposition to obesity. *Nat Genet*. Feb 2018;50(2):180-185. doi:10.1038/s41588-017-0020-9
80. Timshel PN, Thompson JJ, Pers TH. Genetic mapping of etiologic brain cell types for obesity. *eLife*. 2020/09/21 2020;9:e55851. doi:10.7554/eLife.55851
81. Licinio J, Dong C Fau - Wong M-L, Wong ML. Novel sequence variations in the brain-derived neurotrophic factor gene and association with major depression and antidepressant treatment response. (1538-3636 (Electronic))
82. Colle R, Deflesselle E, Martin S, et al. BDNF/TRKB/P75NTR polymorphisms and their consequences on antidepressant efficacy in depressed patients. (1744-8042 (Electronic))

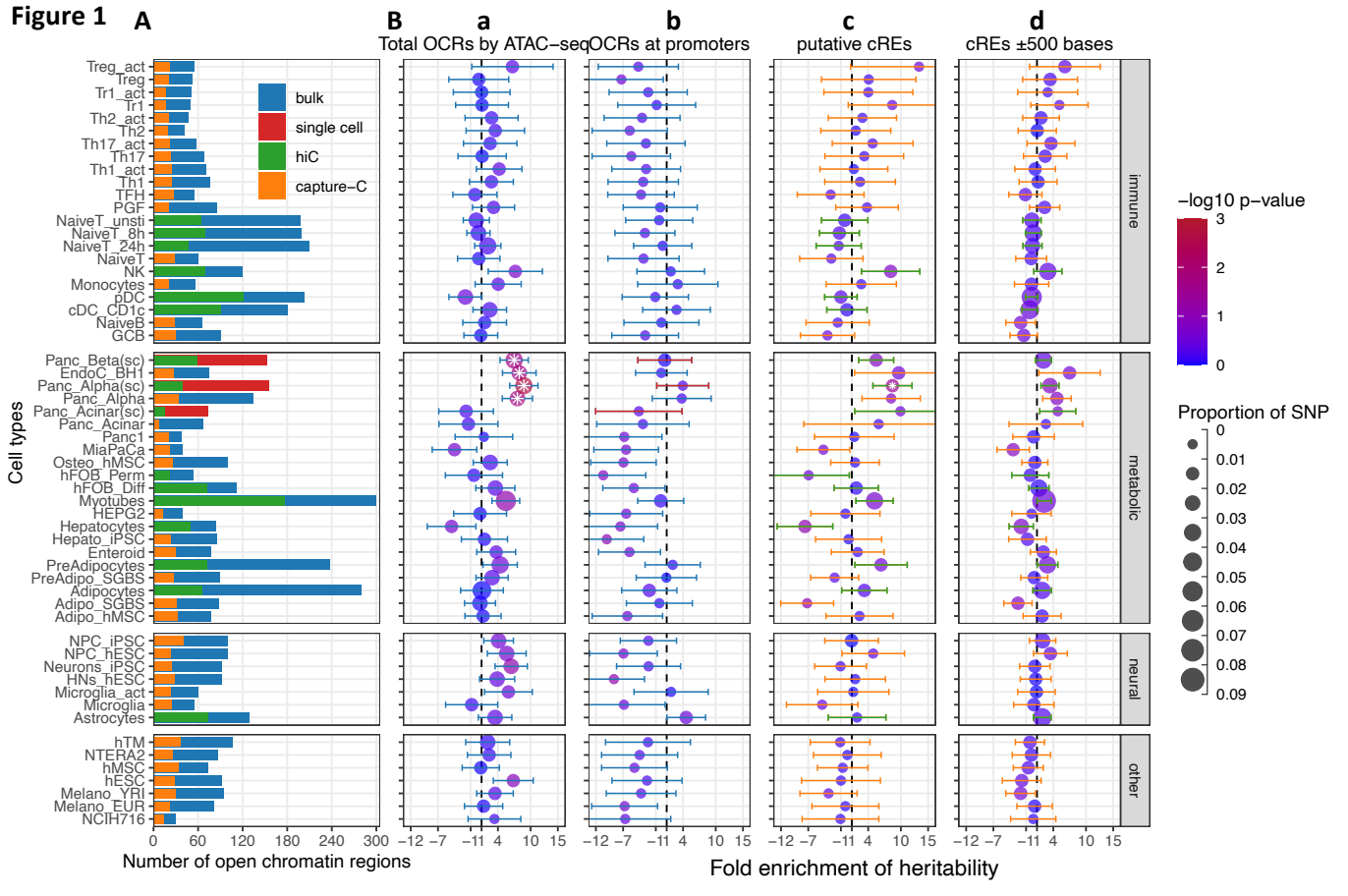


Figure 1
A
B
a
Total OCRs by ATAC−seq
b
OCRs at promoters
c
putative cREs
d
cREs ±500 bases
bulk
single cell
hiC
capture−C
Treg_act
Treg
Tr1_act
Tr1
Th2_act
Th2
Th17_act
Th17
Th1_act
Th1
TFH
PGF
NaiveT_unsti
NaiveT_8h
NaiveT_24h
NaiveT
NK
Monocytes
pDC
cDC_CD1c
NaiveB
GCB
Panc_Beta(sc)
EndoC_BH1
Panc_Alpha(sc)
Panc_Alpha
Panc_Acinar(sc)
Panc_Acinar
Panc1
MiaPaCa
Osteo_hMSC
hFOB_Perm
hFOB_Diff
Myotubes
HEPG2
Hepatocytes
Hepato_iPSC
Enteroid
PreAdipocytes
PreAdipo_SGBS
Adipocytes
Adipo_SGBS
Adipo_hMSC
NPC_iPSC
NPC_hESC
Neurons_iPSC
HNs_hESC
Microglia_act
Microglia
Astrocytes
hTM
NTERA2
hMSC
hESC
Melano_YRI
Melano_EUR
NCIH716
Cell types
immune
metabolic
neural
other
0
60
120
180
240
300
Number of open chromatin regions
−12 −7 −1 1 4 10 15
Fold enrichment of heritability
−log10 p−value
3
2
1
0
Proportion of SNP
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09

Figure 2

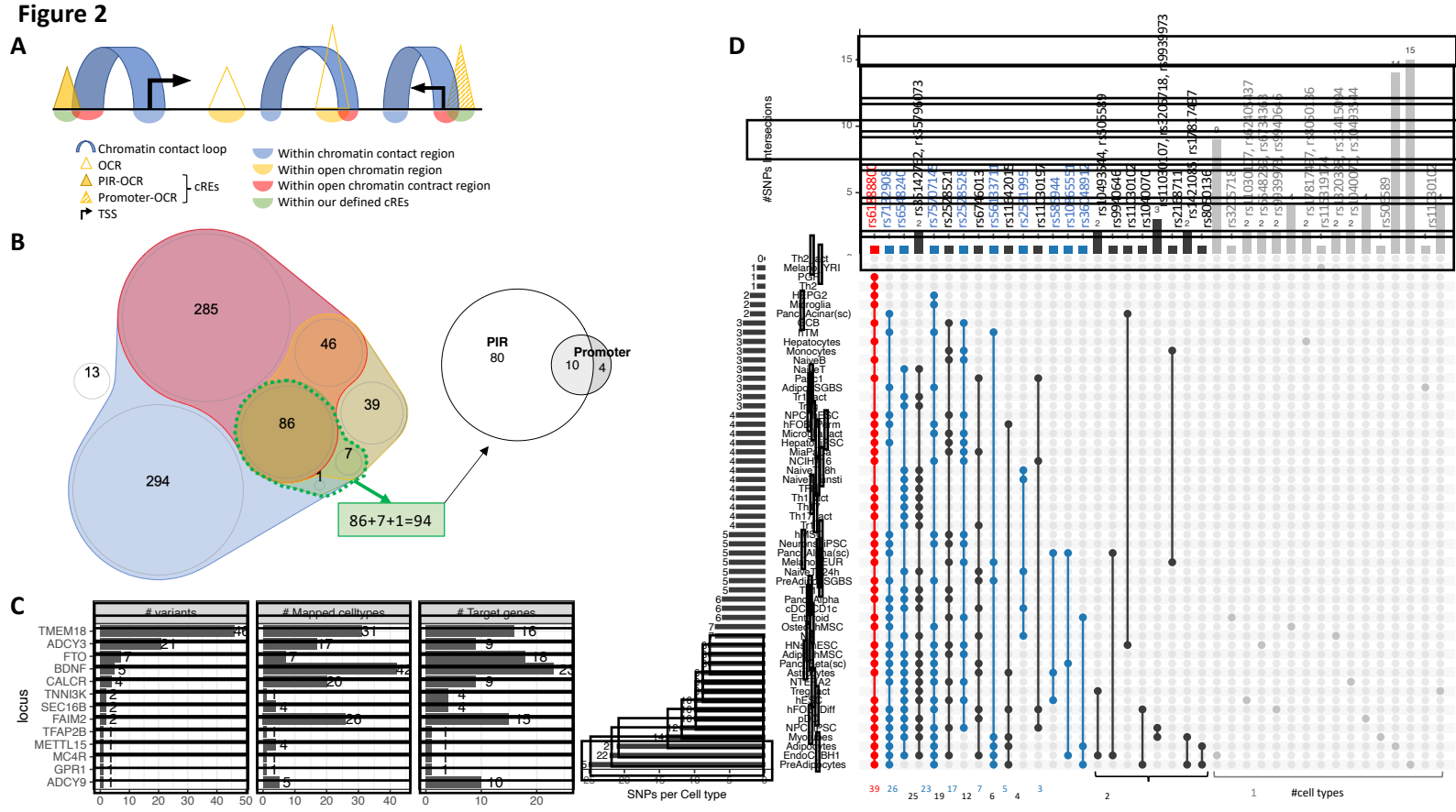

A
Chromatin contact loop
OCR
PIR-OCR
Promoter-OCR
cREs
TSS
Within chromatin contact region
Within open chromatin region
Within open chromatin contract region
Within our defined cREs
B
285
46
13
39
86
7
1
294
PIR
80
Promoter
10
4
86+7+1=94
C
variants
Mapped celltypes
Target genes
locus
TMEM18
ADCY3
FTO
BDNF
CALCR
TNNI3K
SEC16B
FAIM2
TFAP2B
METTL15
MC4R
GPR1
ADCY9
SNPs per Cell type
D
#SNPs intersections
#cell types

Figure 3

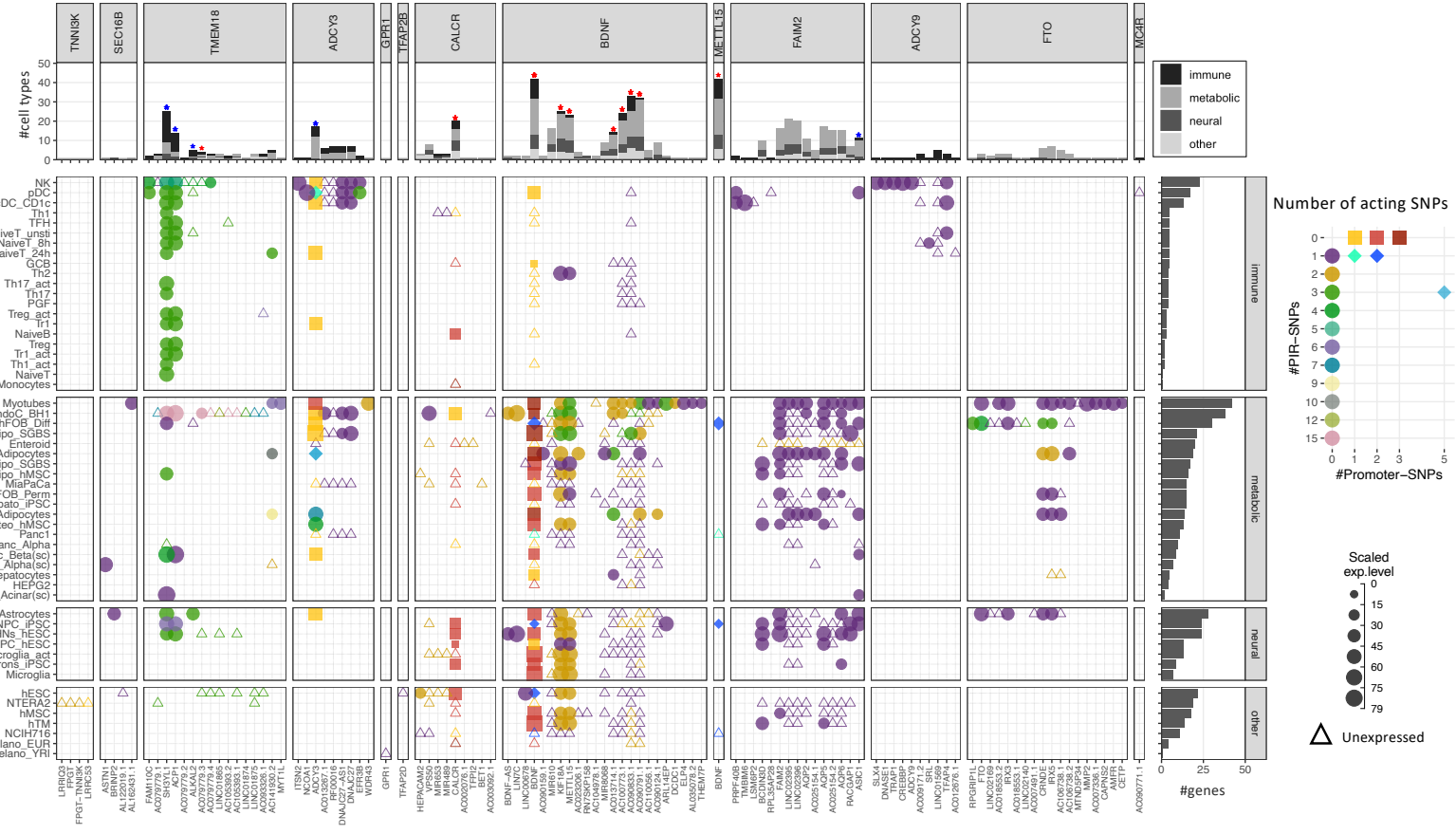
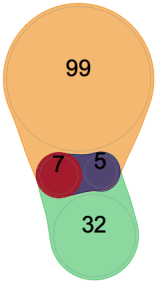

TNNI3K
SEC16B
TMEM18
ADCY3
GPR1
TFAP2B
CALCR
BDNF
METTL15
FAM2
ADCY9
FTO
MC4R
#cell types
immune
metabolic
neural
other
NK
pDC
cDC_CD1c
Th1
TFH
NaiveT_unsti
NaiveT_8h
NaiveT_24h
GCB
Th2
Th17_act
Th17
PGF
Treg_act
Tr1
NaiveB
Treg
Tr1_act
Th1_act
NaiveT
Monocytes
Myotubes
EndoC_BH1
hFOB_Diff
PreAdipo_SGBS
Enteroid
PreAdipocytes
Adipo_SGBS
Adipo_hMSC
MiaPaCa
hFOB_Perm
Hepato_iPSC
Adipocytes
Osteo_hMSC
Panc1
Panc_Alpha
Panc_Beta(sc)
Panc_Alpha(sc)
Hepatocytes
HEPG2
Panc_Acinar(sc)
Astrocytes
NPC_iPSC
HNs_hESC
NPC_hESC
Microglia_act
Neurons_iPSC
Microglia
hESC
NTERA2
hMSC
hTM
NCIH716
Melano_EUR
Melano_YRI
#genes
Number of acting SNPs
#PIR−SNPs
#Promoter−SNPs
Scaled exp.level
Unexpressed

Figure 4

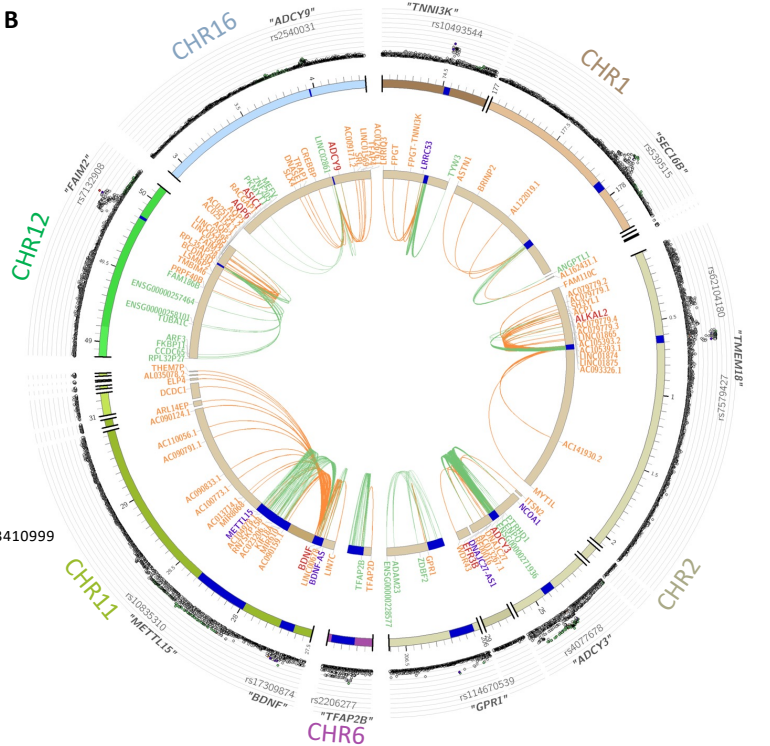
A



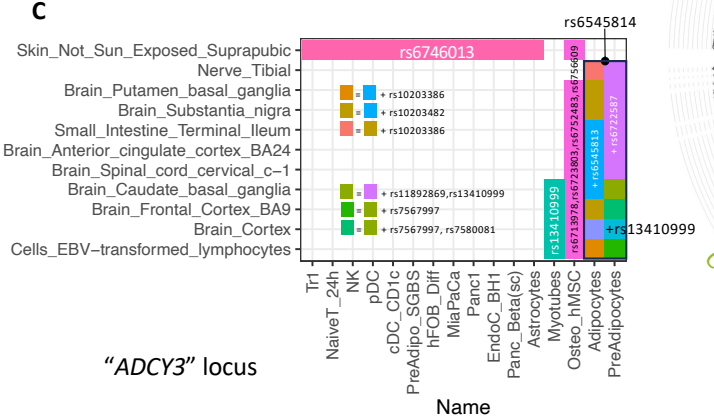
Number of putative target genes identified by:

- ColocQuiaL
- Variant-to-gene mapping
- Overlapped but linked different effector variant
- Overlapped by same effector variant

B



C



"*ADCY3*" locus