1 **Cis-regulatory architecture of human ESC-derived hypothalamic neuron differentiation**

2 **aids in variant-to-gene mapping of relevant complex traits**

3

4 Matthew C. Pahl[1]*, Claudia A. Doege[2]*, Kenyaita M. Hodge[1]*, Sheridan H. Littleton[1]*, Michelle E.

5 Leonard[1], Sumei Lu[1], Rick Rausch[3], James A. Pippin[1], Jonathan P. Bradfield[1], Reza K. Hammond[1], Keith

6 Boehm[1], Robert I. Berkowitz[4], Chiara Lasconi[1,5], Chun Su[1], Alessandra Chesi[1,5], Matthew E. Johnson[1],

7 Andrew D. Wells[1,6,7], Benjamin F. Voight[8,9], Rudolph L. Leibel[10], Diana L. Cousminer[1,5,8]* and Struan F.A.

8 Grant[1,4,5,8]*†

9

10 *equal contribution

11

12 1. Center for Spatial and Functional Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104 USA.

13 2. Department of Pathology, Naomi Berrie Diabetes Center, Columbia Stem Cell Initiative, Columbia University Vagelos College

14   of Physicians and Surgeons, New York, NY, USA.

15 3. Department of Pediatrics, Naomi Berrie Diabetes Center, Columbia Stem Cell Initiative, Columbia University, Vagelos College

16   of Physicians and Surgeons, New York, NY, USA.

17 4. Department of Pediatrics, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104 USA.

18 5. Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104 USA.

19 6. Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

20 7. Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.

21 8. Department of Genetics, University of Pennsylvania, Philadelphia, PA, 19104 USA.

22 9. Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of

23   Pennsylvania, Philadelphia PA, 19104 USA.

24 10. Division of Molecular Genetics (Pediatrics) and the Naomi Berrie Diabetes Center, Columbia University Vagelos College of

25   Physicians and Surgeons, New York, NY, USA.

26

27 †Corresponding author (grants@email.chop.edu)

28

29   **Summary**

30   The hypothalamus regulates metabolic homeostasis by influencing behavior, energy

31   utilization and endocrine systems. Given its role governing health-relevant traits, such as body

32   weight and reproductive timing, understanding the genetic regulation of hypothalamic

33   development and function should yield insights into these traits and diseases. However, given

34   its inaccessibility, studying human hypothalamic gene regulation has proven challenging. To

35   address this gap, we generated a chromatin architecture atlas of an established embryonic stem

36   cell (ESC)-derived hypothalamic-like neuron (HN) model across three stages of *in vitro*

37   differentiation. We profiled accessible chromatin and identified physically interacting contacts

38   between gene promoters and their putative cis-regulatory elements (cREs) to characterize

39   changes in the gene regulatory landscape during hypothalamic differentiation. Next, we

40   integrated these data with GWAS loci for multiple traits and diseases enriched for heritability in

41   these cells, identifying candidate effector genes and cREs impacting transcription factor binding.

42   Our results reveal common target genes for these traits, potentially identifying core

43   hypothalamic developmental pathways. Our atlas will enable future efforts to determine precise

44   mechanisms underlying hypothalamic development with respect to specific disease

45   pathogenesis.

46

47 **INTRODUCTION**

48   The hypothalamus is a critical regulator of many physiological functions, including energy

49 homeostasis, reproduction, sleep and stress[1]. This brain region senses neural and physiological

50 signals, which triggers distinct populations of neurons to release neurotransmitters and peptide

51 neuromodulators to signal the autonomic nervous and endocrine systems[1-3]. Monogenic

52 mutations in key nutrient-sensing hypothalamic genes, such as the leptin and melanocortin 4

53 receptors, result in obesity through dysregulating the neural circuit involved in controlling hunger

54 and satiety, while mutations impacting gonadotrophin-releasing hormone signaling impair the

55 onset of puberty by disrupting pituitary gland signaling[2, 4].

56   There is a lack of epigenomic data characterizing the genetic regulatory architecture of the

57 developing and mature human hypothalamus, limiting our ability to translate studies into

58 information directly relevant for disease[5]. Recently, improvements in embryonic and induced

59 pluripotent stem cell differentiation strategies[1, 6, 7] have partially mitigated the need to study

60 human hypothalamic neurons *ex vivo.* As the precise regulation of hypothalamic development

61 remains poorly understood, differentiating hypothalamic neurons from ESCs provides an

62 opportunity to study these cells and their precursors over time, which should lead to greater

63 understanding of the development of hypothalamic-governed traits and diseases.

64   Genome wide association studies (GWAS) have yielded hundreds of loci statistically

65 associated with phenotypes known to involve hypothalamic function[8-10]. GWAS efforts typically

66 only report single nucleotide polymorphisms (SNPs) yielding the statistically strongest

67 associations per locus. However, these lead SNPs are not necessarily the causal variants due

68 the presence of other SNPs in linkage disequilibrium (LD). The majority of GWAS signals reside

69 in non-coding regions of the genome, suggesting that their impact on phenotype is primarily via

70 gene regulation. As *cis*-acting regulatory elements, such as enhancers or silencers, can act

71 locally or over large genomic distances, the nearest gene to a GWAS signal may not be the

72   principal effector gene[11-14]. Thus, a major challenge in complex trait genetics is to confidently

73   identify the precise regulatory variant(s) tagged by sentinel SNPs and their corresponding

74   effector target gene(s).

75   Chromatin conformation approaches to can identify regions harboring SNPs contacting

76   effector genes via long-range promoter interactions in various cell and tissue contexts[16-18].

77   Recently, we combined a suite of techniques to systematically evaluate GWAS signals located

78   in distal elements[19-22]. Together, our integrated "variant-to-gene mapping" approach aims to

79   physically fine-map significant GWAS loci by identifying open proxy SNPs in LD with each given

80   sentinel signal that directly contact a gene promoter. Assaying relevant cell types in this regard

81   is critical, as promoter architecture varies across cellular identity and developmental-stage[23,16,24].

82   While changes in hypothalamic gene expression during development have been studied [25,

83   [26], the corresponding cis-regulatory architecture in hypothalamus development remains largely

84   unexplored. In this study, we used an arsenal of molecular techniques to characterize the

85   genetic architecture of differentiation of embryonic stem cells, first into hypothalamic progenitors

86   (HPs) and then arcuate (ARC) nucleus-like hypothalamic neurons (HN). We subsequently

87   superimposed GWAS findings for relevant traits on these data to implicate critical and novel

88   effector genes, along with their corresponding putative regulatory elements.

89
90

MATERIALS AND METHODS:

Human ESC-derived hypothalamic neuron differentiation:

The HN differentiation protocol was described previously[27]. Briefly, the human ESC H9 line was seeded on Matrigel plates (16 million cells/148 cm^2; 5 x 148 cm^2 Corning dishes) in ESC medium (KnockOut DMEM supplemented with 15% KnockOut Serum Replacement, 0.1mM MEM Non-Essential Amino Acids, 2mM GlutaMAX, 0.06 mM 2-Mercaptoethanol) with FGF-Basic (AA 1 – 155), (20 ng/ml media) and 10 μM Y-27632. Upon confluency (day 1), cells were cultured in ESC medium without FGF-Basic and Y-27632, but supplemented with Shh (100 ng/ml), Purmorphamine (2 μM), 10 μM SB431542, and 2.5 μM LDN193289. From days 5 to 8, ESC medium was gradually replaced with neuroprogenitor medium (DMEM/F-12 supplemented with 0.1 mM MEM Non-Essential Amino Acids, N-2 Supplement, 0.2 μM ascorbic acid, 0.16% glucose). At day 9, cells were switched into neuronal differentiation medium (DMEM/F-12 supplemented with 0.1 mM MEM Non-Essential Amino Acids, N-2 Supplement, B-27 Supplement minus Vitamin A, 0.2 μM ascorbic acid, 0.16% glucose containing 10 μM DAPT). On day 12, cells were collected with TrypLE Express Enzyme at 37°C for 7 min and washed twice including filtration through pre-wetted 40 μm Corning Sterile Cell Strainer. The hypothalamic progenitor cell pellet was then resuspended with neuronal differentiation medium containing 10 μM Y-27632 for plating on 148 cm^2 dishes coated with Poly-L-Ornithine solution (0.01%) and laminin (4 μg/ml) at a seeding density of 16 million cells/148 cm^2. After 4 hours, the medium was changed to neuronal differentiation medium supplemented with 10 μM DAPT. On day 15, the neuronal differentiation medium was supplemented with 20 ng/ml BDNF until collection on day 27.

ATAC-seq, RNA-seq, Capture C Library generation, Processing, Peak Calling

116 **ATAC-seq library generation, analyses, and peak calling**

117 ATAC-seq library generation:

118   A total of 50,000 cells were centrifuged at 550g for 5 min at 4°C. The cell pellet was

119 washed with cold PBS and resuspended in 50 µL cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10

120 mM NaCl, 3 mM MgCl2, 0.1% NP-40/IGEPAL CA-630) and immediately centrifuged at 550g for

121 10 min at 4°C. Nuclei were resuspended in the Nextera transposition reaction mix (25 µL 2x TD

122 Buffer, 2.5 µL Nextera Tn5 transposase, and 22.5 µL nuclease free H2O) on ice, then incubated

123 for 45 min at 37°C. The tagmented DNA was then purified using the Qiagen MinElute kit and

124 eluted in 10.5 µL Elution Buffer (EB). Ten microliters of purified tagmented DNA was PCR

125 amplified using the Nextera Indexing Kit for 12 cycles to generate each library. The PCR

126 reaction was subsequently purified using 1.8x AMPure XP beads, and concentrations were

127 measured by Qubit Fluorometer. Quality of completed libraries was assessed on a Bioanalyzer

128 2100 high sensitivity DNA Chip. Libraries were paired-end sequenced at the Center for Spatial

129 and Functional Genomics on the Novaseq 6000 platform (50 bp read length).

130

131 Analysis and Peak calling:

132   The number reads from the hypothalamic neurons were down-sampled to make the

133 sequencing depth comparable between conditions using sambamba. Open chromatin regions

134 were called using the ENCODE ATAC-seq pipeline. Pair-end reads from all replicates for each

135 cell type were aligned to the hg19 genome using bowtie2, and duplicate reads were removed

136 from the alignment. Aligned tags were generated by modifying the reads alignment by offsetting

137 +4bp for all the reads aligned to the forward strand, and -5bp for all the reads aligned to the

138 reverse strand. Narrow peaks were called independently for pooled replicates for each cell type

139 using macs2 (-p 0.01 --nomodel --shift -75 --extsize 150 -B --SPMR --keep-dup all --call-

140   summits) and ENCODE blacklist regions were removed from called peaks. We then merged

141   peaks with at least 1 bp overlap between replicates to generate a consensus set of peaks. The

142   consensus set peaks were filtered to those which were reproducible in at least half the ATAC-

143   seq replicates using bedtools intersect[11]. For analyses involving cell-type specific sets of peaks,

144   we considered the set of consensus peaks with mean FPKM value greater than 1 to be "open"

145   in that cell type.

146   For TF analysis replicated, deduplicated ATAC-seq bam files were merge and downsampled to

147   consistent read count for each stage of differentiation to calculate purity scores for each TF.

148   Differential analysis of chromatin accessibility

149   To identify differentially accessible OCRs between ESCs, HPs, and HNs, we used the R

150   package csaw, which uses the de-duplicated read counts for the consensus OCRs for each

151   replicate to normalized against background (10K bins of genome). OCRs with median value of

152   less than 1.2 CPM (~10-50 reads per OCR) across all replicates were removed from further

153   differential analysis. Similar to RNA-seq differential analysis, accessibility differential analysis of

154   the consensus OCRs was performed using glmQLFit approach, fitting cell type in edgeR, but

155   using the csaw normalization scaling factors. Differential OCRs between cell types were

156   identified with thresholds of FDR < 0.05 and absolute log2 fold change > 1. FPKM values were

157   calculated for all OCRs in the consensus list.

158

159   **RNA-seq library generation and analysis:**

160   RNA-seq library generation:

161   RNA was isolated from each cell type in triplicate using TRIzol Reagent. RNA was then

162   purified using the Direct-zol RNA Miniprep Kit and depleted of contaminating genomic DNA

163   using DNAse I. Purified RNA was then checked for quality on the Bioanalyzer 2100 using the

164   Nano RNA Chip and samples with a RIN number above 7 were used for RNA-seq library

165   synthesis. RNA samples were depleted of rRNA using the QIAseq FastSelect RNA Removal Kit

166   then processed into libraries using the NEBNext Ultra II Directional RNA Library Prep Kit for

167   Illumina according to manufacturer's instructions. Quality and quantity of the libraries was

168   measured using the Bioanalyzer 2100 DNA chip and Qubit Fluorometer. Completed libraries

169   were pooled and sequenced on the NovaSeq 6000 platform using paired-end 51bp reads at the

170   Center for Spatial and Functional Genomics at CHOP.

171   RNA-seq processing and differential expression analysis:

172   Sequencing data was demultiplexed and FastQ files were generated using Illumina

173   bcl2fastq2 conversion. Paired-end Fastq files with for each replicate were mapped to the

174   reference genome using STAR. Gene features were assigned to a curated annotation consisting

175   of GencodeV19 with lincRNA and sno/miRNA annotation from the UCSC Table Browser. The

176   raw read count for each gene feature was calculated using HTSeq-count. with parameter

177   settings -f bam -r pos -s reverse -t exon -m intersect. The genes located on chrM or annotated

178   as ribosomal RNAs were removed before further processing.

179   Differential analysis was performed in R using the edgeR package . Briefly, the raw

180   reads per genes features were converted to read Counts Per Million mapped reads (CPM). The

181   gene features with median value of less than 0.7 CPM (10 ~ 18 reads per gene feature) across

182   all samples were filtered. Normalization scaling factors were calculated using the trimmed mean

183   of M-values method. Differentially expressed genes between ESCs, HPs, and HNs were

184   identified with thresholds of FDR < 0.05 and absolute $log_2FC > 1$. Expression values are

185   reported as transcript per million mapped reads (TPM). We clustered standardized TPM values

186   of differentially expressed genes using the R function hclust. Genes expression values were

187   standardized using the R function, scale. Following this, the top six branches were cut to define

188   the clusters used in subsequent comparisons.

189

Capture C Library Generation and Analysis:

3C Library Generation:

We used standard methods for 3C library generation[82]. For each library, 10^7 fixed cells were thawed at 37ºC, followed by centrifugation at RT for 5 mins at 14,000 rpm. The cell pellet was resuspended in 1 mL of dH2O supplemented with 5 μL 200X protease inhibitor cocktail, incubated on ice for 10 mins, then centrifuged. Cell pellet was resuspended to a total volume of 650 μL in dH2O. 50 μL of cell suspension was set aside for pre-digestion QC, and the remaining sample was divided into 3 tubes. Both pre-digestion controls and samples underwent a pre-digestion incubation with the addition of 0.3% SDS, 1x NEBuffer DpnII, and dH2O for 1hr at 37ºC in a Thermomixer shaking at 1,000rpm. A 1.7% solution of Triton X-100 was added to each tube and shaking was continued for another hour. After the pre-digestion incubation, 10 μL of DpnII was added only to each sample tube, and continued shaking along with the pre-digestion control until the end of the day. An additional 10 μL of DpnII was added to each digestion reaction and digestion continued overnight. The next day, another additional 10 μL of DpnII was added and the incubation continued for another 2-3 hours. 100 μL of each digestion reaction was then removed, pooled into one 1.5 mL tube, and set aside for digestion efficiency QC. The remaining samples were heat inactivated at 65°C for 20 min at 1000 rpm in a Thermomixer, and cooled on ice for 20 additional minutes. Digested samples were ligated with 8 μL of T4 DNA ligase and 1X ligase buffer at 1,000 rpm overnight at 16°C in a Thermomixer. The next day, an additional 2 μL of T4 DNA ligase was spiked into each sample and incubated for another few hours. The ligated samples were then de-crosslinked overnight at 65°C with Proteinase K along with the pre-digestion and digestion controls. The following morning, both controls and ligated samples were incubated for 30 min at 37°C with RNase A, followed by phenol/chloroform/isoamyl alcohol (Fisher Cat # BP1752I400) extraction and ethanol precipitation at -20°C. The 3C libraries were centrifuged at 3000 rpm for 45 min at 4°C, while the

215   controls were centrifuged at 14,000 rpm, to pellet the samples. DNA pellets were resuspended

216   in 70% ethanol and again centrifuged as described above. The 3C library pellets and control

217   pellets were resuspended in 300 µL and 20 µL dH2O, respectively, and stored at –20°C.

218   Sample concentrations were measured by Qubit Fluorometer. Digestion and ligation efficiencies

219   were assessed by gel electrophoresis on a 0.9% agarose gel and quantitative PCR (Brilliant III

220   SYBR qPCR Master Mix, VWR Cat # 97066-528).

221

222   Promoter Capture Library Generation:

223   We followed our same protocols as previously published[23]. Isolated DNA from 3C

224   libraries was quantified using a Qubit Fluorometer, and 10 µg of each library was sheared in

225   dH2O using a QSonica Q800R to an average fragment size of 350bp. QSonica settings used

226   were 60% amplitude, 30s on, 30s off, 2 min intervals, for a total of 5 intervals at 4 °C. After

227   shearing, DNA was purified using AMPure XP beads. DNA size was assessed on a Bioanalyzer

228   2100 using a DNA 1000 Chip and DNA concentration was checked via Qubit Fluorometer.

229   SureSelect XT library prep kits were used to repair DNA ends and for adaptor ligation following

230   the manufacturer protocol. Excess adaptors were removed using AMPure XP beads. Size and

231   concentration were checked again by Bioanalyzer 2100 using a DNA 1000 Chip and by Qubit

232   Fluorometer before hybridization. One microgram of adaptor-ligated library was used as input

233   for the SureSelect XT capture kit using manufacturer protocol and our custom-designed 41K

234   promoter Capture-C probe set. The quantity and quality of the captured libraries were assessed

235   by Bioanalyzer using a high sensitivity DNA Chip and by Qubit Fluorometer. SureSelect XT

236   libraries were then paired-end sequenced on Illumina NovaSeq 6000 platform (51bp read

237   length) at the Center for Spatial and Functional Genomics at CHOP.

238

239   Analysis of Capture C:

240   Paired-end reads from three replicates from ESCs, HPs, and HNs were pre-processed

241   using the HICUP pipeline with the default parameters. Reads were aligned to hg19 using

242   bowtie2. We called call significant promoter interactions using the read count from promoters

243   included in our reference bait. As previously reported[23], significant interactions were called using

244   CHiCAGO with default parameters except for bin-size set to 2,500. In addition to our analysis

245   per individual DpnII fragment (1frag), we also called interactions by binning four fragments,

246   which improves detection of long-distance interactions. Significant interactions at 4-DpnII

247   fragment resolution were also called using CHiCAGO. Interactions with a CHiCAGO score > 5 in

248   at least one cell type in either 1-fragment or 4-fragment resolution were considered significant.

249

250   Quality control metrics: Reproducibility between ATAC-seq and RNA-seq samples was

251   determined by principal component analysis and pairwise pearson correlation coefficients

252   between samples. Median expression values were downloaded from GTeX (v7). The spearman

253   rank correlation of the genes with expression level pass a threshold of TPM > 5 in at least one

254   cell/tissue type (16,953 genes) were calculated using Spearman's Correlation Coefficient (cor

255   function in R).

256   Genome tracks were visualized using the python packager pyGenomeTracks version 3.0.

257

258   Genomic Annotations: Promoters were defined as 1500kb upstream and 500kb downstream of

259   the TSS (Genecode V19). Overlapping annotations were assigned to genomic features based

260   on a hierarchy of (1) Promoter, (2) 5'UTR, (3) CDS, (4) 3'UTR, (5) first intron, (6) other introns,

261   or (7) intergenic. The percentage of OCRs overlapping with each feature was visualized as pie

262   charts using ggplot2.

263

264   Genomic coordinate visualization: All coordinates referring to hg19 as the reference genome.

265   **Variant to gene mapping**

266   We identified proxies located in open chromatin and fragments interacting with a bait using

267   bedtools intersect, as performed in our previously published method[18]. We considered all

268   interactions with a proxy SNP located in a distal interaction fragment and those falling within

269   OCRs located in baits. Putative target effector genes were then filtered by expression in each

270   respective cell state (TPM > 1). These genes were functionally annotated by the DAVID

271   functional annotation tool.

272   **Gene set enrichment**:

273   GO and REACTOME datasets annotated in MSigDB (v7.0) were used for gene set

274   enrichment analyses. Statistical significance of gene set enrichment was determined using the

275   hypergeometric test, implemented in the R phyper function.

276

277   **Bioinformatic Analyses:**

278   Transcription factor analysis: PIQ, which integrates TF motif scanning with TF footprinting using

279   DNAase or ATAC-seq data, was used to predict TF binding sites[28]. We scanned JASPAR2020

280   core[29] PWMs against hg19, with ENCODE blacklist regions excluded using the default settings.

281   For downstream analyses we considered TF binding sites passing the default cutoff of purity >

282   0.7. We identified TF motifs enriched in cREs compared the set of nonPIR-OCRs using the R

283   package BiFET, with a cutoff of FDR<0.05,

284

285   Partitioned LD Score Regression: Partitioned heritability was measured using LD Score

286   Regression v1.0.0[30]. Partitioned LDSR requires the GWAS summary statistics and a feature

287   annotation. ESC, HP, and HN annotations were generated using bed files containing positions

288   of the cRE (promoter OCRs + PIR-OCRs) with +/-500bp extension as previously performed[30].

289   We selected a set of traits related to metabolic, endocrine, and neuropsychiatric traits with

290   available GWAS summary statistics **(see Supplemental methods)**.

291

292   Variant to gene mapping: Sentinel SNPs were collected from the most recent large-scale GWAS

293   studies. Proxies for each sentinel were queried using SNiPa using the following parameters:

294   Genome assembly GRCh37; Variant set 1000 Genomes, Phase 3 v5; Population European;

295   Genome annotation Ensembl 87 and r^2 > 0.6. Intersection as done previously (see

296   Supplemental Methods). Analysis was then restricted to expressed genes with a TPM > 1.

297

298   GWAS Colocalization: Summary statistics for 6 regions with overlapping associations for 3-4

299   input traits were imputed using FIZI. Common variants (MAF ≥ 0.01) from the European

300   ancestry 1000 Genomes Project v3 samples were used as a reference panel for the imputation.

301   Default parameters were used with the exception that the minimum proportion parameter was

302   lowered to 0.01. Standard errors and betas for the imputed SNPs were estimated using the

303   method from https://github.com/zkutalik/ssimp_software/blob/master/extra/transform_z_to_b.R.

304   Subsequently, HyPrColoc was used to test for colocalization across all input traits

305   simultaneously. Separately, we tested for colocalization for each input trait genome-wide against

306   GTEx v.7 hypothalamic eQTLs using coloc[31].

307

308   **RESULTS**

309

310   ESC-derived hypothalamic-like neurons (HN) recapitulate molecular characteristics of the

311   hypothalamus

312   We utilized an established protocol to derive ARC HN-like neurons from a human ESC

313   line (H9)[32, 33], and collected cells at three stages of differentiation: pluripotent ESCs, NKX2-1+

314   hypothalamic progenitors (HPs), and HNs. We then profiled global gene expression patterns for

315   these three stages using RNA-seq, chromatin accessibility with ATAC-seq, and chromatin

316   conformation via promoter-focused Capture C to generate a high-resolution atlas of the distal

317   promoter interaction landscape in an *in vitro* human model of hypothalamic development

318   (**Figure 1A**). To assess the reproducibility between replicates, we performed principal

319   component analysis and pairwise Pearson correlation on the RNA-seq and ATAC-seq datasets.

320   In both cases the first principal component corresponded to the stage of differentiation and

321   accounted for more than half of the variation (RNA-seq: 52.60%; ATAC-seq: 55.30%)

322   (**Supplemental Figure 1A-D**).

323   To further confirm the molecular congruence of HN differentiation to the *in vivo*

324   development of HNs, we examined the expression of several marker genes (**Figure 1B;**

325   **Supplemental Figure 1E**)[25], which were consistent with expectations [27, 32, 33]. In particular, we

326   investigated the expression and promoter interaction landscape of *NKX2-1*, which encodes a

327   transcription factor (TF) critical for hypothalamus specification. *NKX2-1* is expressed in the

328   developing hypothalamus and subsequently becomes restricted to a subset of neurons[34]. *NKX2-*

329   *1* expression followed a similar pattern during HN differentiation (**Figure 1B**). In addition, we

330   observed a distinct change in the accessibility of the *NKX2-1* promoter, concordant with its

331   expression pattern (**Figure 1C**), as well as fewer interactions detected as *NKX2-1* expression

332   decreased (**Figure 1D**), confirming our detection of expected dynamic changes.

333   In addition to confirming the expression of known marker genes, we compared the global

334   transcriptomic profile of HNs to the GTeX RNA-seq database, which is derived from primary

335   human tissue samples (**Figure 1E; Supplemental Table 1**)[35]. HN gene expression was highly

336   correlated with the hypothalamus (Spearman's rho= 0.719; adjusted P = 1.14 x 10^{-14}). Taken

337   together, these results show that ESC-derived HNs resemble hypothalamus tissue.

338

339   *Temporal dynamics of regulation of gene expression and cis-regulatory elements during*

340   *hypothalamic neuron differentiation*

341   We assessed the temporal profile of gene expression to identify genes with developmental

342   stage-restricted expression during human HN differentiation, with 15,808 genes differentially

343   expressed in at least one stage **(Figure 2A).** We assigned these genes to six clusters based on

344   expression patterns during the course of differentiation. Each cluster corresponded to genes

345   specifically enriched or depleted in at least one stage of differentiation (**Figure 2B;**

346   **Supplemental Figure 2A-B**). Gene Ontology (GO) and REACTOME enrichment analysis of

347   each cluster identified known biological processes related to proliferating progenitor cells and

348   differentiated neurons[36] (**Supplemental Figure 2 C-D**).

349   To correlate gene expression changes during HN differentiation with the respective

350   chromatin accessibility and conformation profiles at each stage, we defined the relationship with

351   open chromatin regions (OCRs) using ATAC-seq. We identified a total of 404,691 OCRs in at

352   least one stage. The OCRs were disproportionately located in promoters (-1500/+500bp TSS)

353   and first introns **(Supplemental Figure 3A)**, which is characteristic of regulatory elements[37]. We

354   then grouped the OCRs into three categories **(Figure 2C)**: (1) OCRs located within promoter

355   regions annotated as 'promoter OCRs'; (2) OCRs with direct promoter contacts determined by

356   Capture C, annotated as 'promoter-interacting region (PIR)-OCRs'; and (3) OCRs that could not

357   be assigned to a gene because they did not fit either criteria, annotated as 'nonPIR-OCRs'.

358   Because they could be annotated to a gene, we considered the sets of 50,952 promoter OCRs

359   and 87,170 PIR-OCRs as putative cREs.

360   Both the number of cREs per gene and the mean distance between the cRE and the

361   promoter were decreased in HPs compared to ESCs or HNs **(Supplemental Figure 3B-C)**,

362   reflecting fewer long-range interactions detected at this stage **(Supplemental Figure 3D-E)**.

363   We also observed a trend for genes with higher expression interacting with more cREs (Kruskal

364   Wallis test: p-value < 2.2 x10^{-16}) **(Supplemental Figure 3F)**, which is in line with reports for

365   other neuronal[38] and immune cells[22].

366   We then compared chromatin accessibility across the three stages, and identified 87,761

367   differentially accessible regions from ESCs to HPs (43,170 more closed; 44,591 more open) and

368   48,522 differentially accessible regions from HPs to HNs (33,642 more closed; 14,880 more

369   open) **(Figure 2G; Supplemental Figure 3G)**. The genome-wide decrease in open chromatin

370   as cells advanced in differentiation occurs in other developing tissues[39]. The subset of cREs

371   contacting differentially expressed genes showed a net increase in accessibility as ESCs

372   differentiated into HPs, and subsequently decreased as HPs advanced to HNs, regardless of

373   gene expression pattern defined by our clustering analysis (**Figure 2H; Figure 2A**). This trend

374   suggests that regulatory elements driving the gene expression changes specifying

375   undifferentiated progenitors to a hypothalamic cell fate are primarily first established by opening

376   of selected cREs followed by a more global pruning of contacts upon differentiation to neurons.

377

378   *Predicting transcription factors controlling HN development from spatial gene regulatory*

379   *architecture*

380   TFs regulate gene expression by binding to specific DNA sequences such as enhancers

381   and silencers. Local chromatin accessibility is a critical determinant of where and when TFs bind

382   to DNA[40]. To identify TFs that may bind to cREs, we leveraged PIQ, which uses chromatin

383   accessibility profiles to improve motif score-based matching [28]. We identified putative binding

384   sites in each stage of differentiation and observed that more binding sites were detected in HNs

385   compared to ESCs or HPs (**Figure 3A)**. After grouping TFs by family, we detected more binding

386   sites for Homeodomain TF factors in HNs compared to ESCs or HPs (**Figure 3B)**. This result

387   was expected, as neuronal identity is refined by the expression of multiple patterning genes[41].

388   We also observed fewer AP-1 family binding sites in HNs compared to ESCs and HPs; these

389   TFs regulate the cell cycle in early cellular development[42].

390   Next, we checked for TF enrichment in cREs in each cell type to identify which TFs could

391   mediate these promoter contacts (**Figure 2C**). We compared the three stages of differentiation

392   for enriched binding in the cREs compared to non-PIR OCRs. This approach generated a set of

393   potentially relevant TFs involved in HN differentiation. We found 474 enriched TFs in ESCs, 122

394   in HPs and 134 in HNs (**Figure 3C-E; Supplemental Table 2**). While some TFs involved in

395   DNA looping, such as MAZ and CTCF, were enriched in all three cell types, we also observed

396   differences in the top enriched TFs in each comparison such as ZBTB6, EBF2, EGR1, and

397   ZIC/MYC family members (**Supplemental Figure 4**).

398

399   *GWAS loci enriched in hypothalamus cREs*

400   We hypothesized that cREs in our hypothalamic model are enriched for genetic variants

401   associated with traits that are at least partly governed by the hypothalamus. We used

402   Partitioned Linkage Disequilibrium Score Regression (LDSR)[30] to identify significantly enriched

403   traits for associated loci falling into hypothalamic cREs. We assembled GWAS summary

404   statistics from several recent studies examining metabolic, circadian, neuropsychiatric and

405   puberty-relevant phenotypes, and tested whether cREs were enriched for GWAS loci in at least

406   one stage of differentiation (**Figure 4A**). We detected significantly enriched signals with BMI,

407   adult height, age at menarche (AAM), major depressive disorder (MDD), bipolar disorder,

408   several measures of sleep (**Figure 4A; Supplemental Table 3**). The enrichment of GWAS loci

409   for these hypothalamus-related traits in the cREs identified in the ESC-derived hypothalamic cell

410   types highlights their utility as a model for gaining insight into the target effector genes and

411   regulatory elements functionally related to these traits and diseases.

412

413   _Variant-to-gene mapping identifies target effector genes at GWAS-implicated loci_

414   Guided by the results of the partitioned LDSR analyses, we performed variant-to-gene

415   mapping for those traits that displayed significant heritability enrichment in at least one of the

416   three cellular differentiation stages. We began with all genome-wide significant loci in the most

417   recent large-scale GWAS for each respective trait and queried for proxy SNPs in LD with each

418   sentinel SNP. We overlapped this set of SNPs with the open chromatin regions identified by

419   ATAC-seq, and queried our promoter focused Capture C data to determine the genes in

420   physical contact with open proxy SNPs in each of the three cell states. Finally, we filtered by

421   expression from our RNA-seq data to limit subsequent analyses to genes expressed in at least

422   one stage of differentiation (TPM > 1) (**Table 1**).

423   For each trait, we noticed that multiple contacted genes also have previously characterized

424   relevant monogenic disease mutations, suggesting that our approach can identify genes with

425   known mechanistic links to the queried traits (**Supplemental Table 4**). For BMI, we detected

426   genes that are known to influence monogenic forms of extreme body weight, including

427   _ABCC8_[44], _BDNF_[45], and _PPARG_[46]. From a AAM locus, we observed _FEZF1,_ known to harbor

428   monogenic mutations that cause delayed/absent puberty[47]. Finally, among sleep traits, our data

429   implicated _PER2_, which encodes a factor that plays a role in advanced sleep phase syndrome[48].

430   Many additional putative effector genes also have plausible biological links to each trait, while

431   others represent novel findings in the context of these phenotypes (**Supplemental Table 5**).

432

433   *Variant contacted genes in hypothalamic development*

434   To identify variants impacting hypothalamus development, we compared our contacted

435   SNPs with the set of cREs connected to specific genes for each trait, and found that ≥50% of

436   the SNP contacted genes were differentially expressed during HN differentiation (**Figure 4B**).

437   To identify biological functions associated with these genes, we tested for GO term enrichment

438   specific to either HP or HNs. HPs were enriched for ERK1/ERK2 cascade and phospho-inositol

439   3 lipid signaling (**Supplemental Table 6**), which are known regulators of neural stem cell

440   proliferation[50-52]. HNs were enriched for clathrin-dependent endocytosis and IRE1-mediated

441   unfolded protein response (**Supplemental Table 6**). Endocytosis is critical for neuronal vesicle

442   recycling at synapses and endoplasmic reticulum stress affects the response of the

443   hypothalamus to external stimuli in obesity[53]. Thus, pathway analysis confirms that the

444   implicated effector genes are likely to be important for hypothalamic development and function.

445

446   *Colocalization of loci associated with multiple traits*

447   Next, we identified contacted genes implicated in the context of multiple GWAS traits.

448   Although most implicated genes were specific to individual traits, we identified multiple genes

449   that were shared, suggesting a degree of overlap in the regulatory mechanisms controlling

450   these traits **(Figure 4C; Supplemental Table 7)**. In particular, two loci contacted four genes

451   (*BSN/FAM212A* and *FEZF1 /FEZF1-AS1*) which were identified in our scans of BMI, height,

452   AAM and sleep. To determine whether these overlaps represent likely shared regulatory

453   regions, we performed Hypothesis Prioritization in multi-trait Colocalization (HyPrColoc) analysis

454   for several regions. Our results highlighted both shared and distinct regulatory architectures

455   across traits that varied by locus. For example, the *FEZF1* region colocalized among BMI,

456   height, and AAM (regional posterior probability (PP) = 0.91), indicating a likely shared regulatory

457   region impacting each trait (**Figure 4D**). Interestingly, the proxy of the *FEZF1* signal was located

458   in a putative NRF1 binding site (**Figure 4D; Supplemental Table 8**). In contrast to the *FEZF1*

459   locus, although the well-known known *BDNF* was implicated as an effector gene for AAM, BMI,

460   and sleep, these three signals appeared to be distinct (PP = 0), suggesting a complex

461   regulatory architecture for this region that differs by trait (**Figure 4E**).

462

463   *Colocalization of target effector genes with eQTLs – cumulative evidence*

464   Multiple data sources can contribute orthogonal evidence for effector genes at GWAS loci[49].

465   The GTEx consortium[35] has characterized hypothalamic tissue eQTLs, so we performed

466   colocalization analyses to assess how many gene-SNP connections agreed with the physical

467   variant-to-gene mapping approach in our specific cellular settings. For AAM, 13 genes

468   colocalized with eQTLs, with two adjacent genes supported by our variant-to-gene mapping

469   approach, *RPS26* (PP = 0.951) and *SUOX* (PP = 0.942). For BMI, we observed 12 colocalized

470   genes, with one gene supported by our variant-to-gene mapping approach, *DHRS11* (PP =

471   0.822). Of the 29 genes colocalized with eQTLs for height, three were supported by our data:

472   *NMT1* (PP = 0.94), *RFT1* (PP = 0.85) and *RPS9* (PP = 0.75). There was only one eQTL

473   colocalized for sleep, but was not detected by our approach. For MDD and bipolar disorder, no

474   genes colocalized with the eQTL data, which may be due to the relatively few signals detected

475   in the eQTL analysis.

476

477

478   **DISCUSSION**

479   We used an established *in vitro* HN model in order to both understand its genomic

480   architecture and to gain insight into mechanisms by which non-coding GWAS loci associated

481   with hypothalamic-regulated traits could mediate their effects. Given the challenge in acquiring

482   primary human hypothalamic tissue and the organ's complex makeup of cell and neuronal

483   types[54], we leveraged RNA-seq, ATAC-seq and promoter-focused Capture C to identify an

484   aggregation of potentially relevant cis-regulatory regions in ESC-derived HNs. Importantly, we

485   verified that HNs exhibited temporal transcriptional profiles that are congruent with *in vivo*

486   hypothalamic molecular expression signatures and functional networks[27].

487   We mapped common GWAS variants associated with AAM, BMI, height, bipolar disorder,

488   sleep, and MDD to putative effector genes via their likely cREs. This approach identified both

489   known and novel genes. For example, *FEZF1* mutations cause hypogonadotropic

490   hypogonadism with anosmia[47]. *FEZF1* is a zinc finger transcriptional repressor that is critical for

491   hypothalamus development[55]. The proxy contacting the *FEZF1* promoter is located in a binding

492   site for NRF1, a transcription factor that regulates expression of several genes involved in

493   mitochondrial biosynthesis and respiration, but is also important for neuronal differentiation and

494   axogenesis[56]. *FEZF1* mutations impair puberty by disrupting the migration of gonadotropin-

495   releasing hormone neurons, which are necessary to initiate puberty, from the olfactory bulb

496   placode to the hypothalamus during fetal development[47]. In contrast, while *BDNF* was

497   implicated in three traits, we observed distinct GWAS association landscapes, with different

498   sentinels pointing to different proxies that consistently contacted the *BDNF* promoter. Thus,

499   *BDNF* appears to have an intricate regulatory architecture and harbors multiple trait-associated

500   variants that likely act in cell-type and temporally-specific contexts.

501   To uncover genes implicated by multiple analytic approaches, we also performed

502   colocalization analyses of the implicated traits with hypothalamic eQTLs. Both eQTL and

503   variant-to-gene mapping approaches identified *DHRS11* for BMI. The overlap between the two

504   approaches was low, possibly due to differences between *ex vivo* tissue samples and stem cell-

505   derived cells. A confluence of evidence is critical for distinguishing true effector genes from the

506   many 'bystander genes' identified in eQTL studies[49]; our physical variant-to-gene mapping

507   pipeline represents one such approach.

508   Here we report aspects of genomic architecture of a stem cell-based model of human

509   hypothalamic development. We relate this architecture to the cellular ontogenesis of the human

510   hypothalamus, and to the regulation of genes that influence complex phenotypes. Application of

511   these strategies enables specific gene attributions for non-coding SNPs implicated in relevant

512   common traits by GWAS efforts. These integrated datasets therefore offer valuable insight for

513   prioritizing candidate genes that drive the molecular mechanisms by which the hypothalamus

514   contributes to the pathogenesis of relevant complex traits.

515

526

527 **DECLARATION OF INTERESTS**

528 The authors declare no conflicts of interest.

REFERENCES

1. Merkle FT, Maroof A, Wataya T, Sasai Y, Studer L, Eggan K *et al.* Generation of neuropeptidergic hypothalamic neurons from human pluripotent stem cells. *Development* 2015; **142**(4)**:** 633-643.

2. Andermann ML, Lowell BB. Toward a Wiring Diagram Understanding of Appetite Control. *Neuron* 2017; **95**(4)**:** 757-778.

3. Yoo S, Blackshaw S. Regulation and function of neurogenesis in the adult mammalian hypothalamus. *Prog Neurobiol* 2018; **170:** 53-66.

4. Herbison AE. Control of puberty onset and fertility by gonadotropin-releasing hormone neurons. *Nat Rev Endocrinol* 2016; **12**(8)**:** 452-466.

5. Rajamani U, Gross AR, Hjelm BE, Sequeira A, Vawter MP, Tang J *et al.* Super-Obese Patient-Derived iPSC Hypothalamic Neurons Exhibit Obesogenic Signatures and Hormone Responses. *Cell Stem Cell* 2018; **22**(5)**:** 698-712.e699.

6. Rajamani U, Gross AR, Hjelm BE, Sequeira A, Vawter MP, Tang J *et al.* Super-Obese Patient-Derived iPSC Hypothalamic Neurons Exhibit Obesogenic Signatures and Hormone Responses. *Cell Stem Cell* 2018; **22**(5)**:** 698-712 e699.

7. Wang L, Egli D, Leibel RL. Efficient Generation of Hypothalamic Neurons from Human Pluripotent Stem Cells. *Curr Protoc Hum Genet* 2016; **90:** 21 25 21-21 25 14.

8. Dashti HS, Jones SE, Wood AR, Lane JM, van Hees VT, Wang H *et al.* Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat Commun* 2019; **10**(1)**:** 1100.

9. Jones SE, Lane JM, Wood AR, van Hees VT, Tyrrell J, Beaumont RN *et al.* Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat Commun* 2019; **10**(1)**:** 343.

10. Jones SE, van Hees VT, Mazzotti DR, Marques-Vidal P, Sabia S, van der Spek A *et al.* Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. *Nat Commun* 2019; **10**(1)**:** 1585.

566 11. Wang L, De Solis AJ, Goffer Y, Birkenbach KE, Engle SE, Tanis R *et al.* Ciliary gene
567   RPGRIP1L is required for hypothalamic arcuate neuron development. *JCI Insight* 2019;
568   **4**(3).
569

570 12. Stratigopoulos G, Burnett LC, Rausch R, Gill R, Penn DB, Skowronski AA *et al.*
571   Hypomorphism of Fto and Rpgrip1l causes obesity in mice. *J Clin Invest* 2016; **126**(5)**:**
572   1897-1910.
573

574 13. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C *et al.* FTO
575   Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med* 2015;
576   **373**(10)**:** 895-907.
577

578 14. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C *et al.* Obesity-
579   associated variants within FTO form long-range functional connections with IRX3. *Nature*
580   2014; **507**(7492)**:** 371-375.
581

582 15. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM *et al.* A
583   common variant in the FTO gene is associated with body mass index and predisposes to
584   childhood and adult obesity. *Science* 2007; **316**(5826)**:** 889-894.
585

586 16. Siersbaek R, Madsen JGS, Javierre BM, Nielsen R, Bagge EK, Cairns J *et al.* Dynamic
587   Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. *Mol*
588   *Cell* 2017; **66**(3)**:** 420-435 e425.
589

590 17. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S *et al.* Lineage-
591   Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to
592   Target Gene Promoters. *Cell* 2016; **167**(5)**:** 1369-1384 e1319.
593

594 18. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL *et al.* A Compendium of Chromatin
595   Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Rep* 2016;
596   **17**(8)**:** 2042-2059.
597

598 19. Chesi A, Wagley Y, Johnson ME, Manduchi E, Su C, Lu S *et al.* Genome-scale Capture
599   C promoter interactions implicate effector genes at GWAS loci for bone mineral density.
600   *Nat Commun* 2019; **10**(1)**:** 1260.
601

602 20. Caliskan M, Manduchi E, Rao HS, Segert JA, Beltrame MH, Trizzino M *et al.* Genetic
603   and Epigenetic Fine Mapping of Complex Trait Associated Loci in the Human Liver. *Am*
604   *J Hum Genet* 2019; **105**(1)**:** 89-107.
605

606   21.   Cousminer DL, Wagley Y, Pippin JA, Elhakeem A, Way GP, McCormack SE *et al.*
607   Genome-wide association study implicates novel loci and reveals candidate effector
608   genes for longitudinal pediatric bone accrual through variant-to-gene mapping. *medRxiv*
609   2020.
610

611   22.   Su C, Johnson ME, Torres A, Thomas RM, Manduchi E, Sharma P *et al.* Human
612   follicular helper T cell promoter connectomes reveal novel genes and regulatory
613   elements at SLE GWAS loci. *bioRxiv* 2019.
614

615   23.   Jung I, Schmitt A, Diao Y, Lee AJ, Liu T, Yang D *et al.* A compendium of promoter-
616   centered long-range chromatin interactions in the human genome. *Nat Genet* 2019;
617   **51**(10)**:** 1442-1449.
618

619   24.   Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y *et al.*
620   Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 2017; **171**(3)**:**
621   557-572.e524.
622

623   25.   Shimogori T, Lee DA, Miranda-Angulo A, Yang Y, Wang H, Jiang L *et al.* A genomic
624   atlas of mouse hypothalamic development. *Nat Neurosci* 2010; **13**(6)**:** 767-775.
625

626   26.   Huisman C, Cho H, Brock O, Lim SJ, Youn SM, Park Y *et al.* Single cell transcriptome
627   analysis of developing arcuate nucleus neurons uncovers their key developmental
628   regulators. *Nat Commun* 2019; **10**(1)**:** 3696.
629

630   27.   Wang L, Meece K, Williams DJ, Lo KA, Zimmer M, Heinrich G *et al.* Differentiation of
631   hypothalamic-like neurons from human pluripotent stem cells. *J Clin Invest* 2015; **125**(2)**:**
632   796-808.
633

634   28.   Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP *et al.*
635   Discovery of directional and nondirectional pioneer transcription factors by modeling
636   DNase profile magnitude and shape. *Nat Biotechnol* 2014; **32**(2)**:** 171-178.
637

638   29.   Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA *et al.*
639   JASPAR 2020: update of the open-access database of transcription factor binding
640   profiles. *Nucleic Acids Res* 2020; **48**(D1)**:** D87-D92.
641

642   30.   Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR *et al.* Partitioning
643   heritability by functional annotation using genome-wide association summary statistics.
644   *Nat Genet* 2015; **47**(11)**:** 1228-1235.
645

646   31.   Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C *et al.*
647   Bayesian test for colocalisation between pairs of genetic association studies using
648   summary statistics. *PLoS Genet* 2014; **10**(5)**:** e1004383.
649

650   32.   Burnett LC, Leduc CA, Sulsona CR, Paull D, Rausch R, Eddiry S *et al.* Deficiency in
651   prohormone convertase PC1 impairs prohormone processing in Prader-Willi syndrome.
652   *Journal of Clinical Investigation* 2016; **127**(1)**:** 293-305.
653

654   33.   Burnett LC, LeDuc CA, Sulsona CR, Paull D, Rausch R, Eddiry S *et al.* Deficiency in
655   prohormone convertase PC1 impairs prohormone processing in Prader-Willi syndrome. *J
656   Clin Invest* 2017; **127**(1)**:** 293-305.
657

658   34.   Yee CL, Wang Y, Anderson S, Ekker M, Rubenstein JL. Arcuate nucleus expression of
659   NKX2.1 and DLX and lineages expressing these transcription factors in neuropeptide
660   Y(+), proopiomelanocortin(+), and tyrosine hydroxylase(+) neurons in neonatal and adult
661   mice. *J Comp Neurol* 2009; **517**(1)**:** 37-50.
662

663   35.   Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical
664   Methods groups-Analysis Working G, Enhancing Gg, Fund NIHC *et al.* Genetic effects
665   on gene expression across human tissues. *Nature* 2017; **550**(7675)**:** 204-213.
666

667   36.   Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA *et al.* Gene
668   set enrichment analysis: a knowledge-based approach for interpreting genome-wide
669   expression profiles. *Proc Natl Acad Sci U S A* 2005; **102**(43)**:** 15545-15550.
670

671   37.   Park SG, Hannenhalli S, Choi SS. Conservation in first introns is positively associated
672   with the number of exons within genes and the presence of regulatory epigenetic
673   signals. *BMC Genomics* 2014; **15:** 526.
674

675   38.   Song M, Yang X, Ren X, Maliskova L, Li B, Jones IR *et al.* Mapping cis-regulatory
676   chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target
677   genes. *Nat Genet* 2019; **51**(8)**:** 1252-1262.
678

679   39.   Ma Y, McKay DJ, Buttitta L. Changes in chromatin accessibility ensure robust cell cycle
680   exit in terminally differentiated cells. *PLoS Biol* 2019; **17**(9)**:** e3000378.
681

682   40.   Slattery M, Zhou T, Yang L, Dantas Machado AC, Gordan R, Rohs R. Absence of a
683   simple code: how transcription factors read the genome. *Trends Biochem Sci* 2014;
684   **39**(9)**:** 381-399.
685

41. Briscoe J, Pierani A, Jessell TM, Ericson J. A Homeodomain Protein Code Specifies Progenitor Cell Identity and Neuronal Fate in the Ventral Neural Tube. *Cell* 2000; **101**(4)**:** 435-445.

42. Hilger-Eversheim K. MM, Schorle H., Buettner R. Regulatory roles of AP-2 transcription factors in vertebrate development, apoptosis, and cell cycle control. *Genes Cells* 2000; **260:** 1-12.

43. Luo Z, Gao X, Lin C, Smith ER, Marshall SA, Swanson SK *et al.* Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Mol Cell* 2015; **57**(4)**:** 685-694.

44. Leslie J. Baier YLM, Maria Sara Remedi, Michael Traurig,, Paolo Piaggi GW, Ke Huang, Alyssa Stacy, Sayuko Kobes,, Jonathan Krakoff PHB, Robert G. Nelson, William C. Knowler,, Robert L. Hanson CGN, and Clifton Bogardus. ABCC8 R1420H Loss-of-Function Variant in a Southwest American Indian Community: Association With Increased Birth Weight and Doubled Risk of Type 2 Diabetes. *Diabetes* 2015; **64:** 4322-4332.

45. Sandrini L, Di Minno A, Amadio P, Ieraci A, Tremoli E, Barbieri SS. Association between Obesity and Circulating Brain-Derived Neurotrophic Factor (BDNF) Levels: Systematic Review of Literature and Meta-Analysis. *Int J Mol Sci* 2018; **19**(8).

46. Moller DE, Berger JP. Role of PPARs in the regulation of obesity-related insulin sensitivity and inflammation. *Int J Obes Relat Metab Disord* 2003; **27 Suppl 3:** S17-21.

47. Kotan LD, Hutchins BI, Ozkan Y, Demirel F, Stoner H, Cheng PJ *et al.* Mutations in FEZF1 cause Kallmann syndrome. *Am J Hum Genet* 2014; **95**(3)**:** 326-331.

48. Miyagawa T, Hida A, Shimada M, Uehara C, Nishino Y, Kadotani H *et al.* A missense variant in PER2 is associated with delayed sleep-wake phase disorder in a Japanese population. *J Hum Genet* 2019; **64**(12)**:** 1219-1225.

49. Ndungu A, Payne A, Torres JM, van de Bunt M, McCarthy MI. A Multi-tissue Transcriptome Analysis of Human Metabolites Guides Interpretability of Associations Based on Multi-SNP Models for Gene Expression. *Am J Hum Genet* 2020; **106**(2)**:** 188-201.

724   50.   Sato A, Sunayama J, Matsuda K, Tachibana K, Sakurada K, Tomiyama A *et al.*
725   Regulation of neural stem/progenitor cell maintenance by PI3K and mTOR. *Neurosci*
726   *Lett* 2010; **470**(2)**:** 115-120.
727

728   51.   Imamura O, Pages G, Pouyssegur J, Endo S, Takishima K. ERK1 and ERK2 are
729   required for radial glial maintenance and cortical lamination. *Genes Cells* 2010; **15**(10)**:**
730   1072-1088.
731

732   52.   Le Belle JE, Orozco NM, Paucar AA, Saxe JP, Mottahedeh J, Pyle AD *et al.* Proliferative
733   neural stem cells have high endogenous ROS levels that regulate self-renewal and
734   neurogenesis in a PI3K/Akt-dependant manner. *Cell Stem Cell* 2011; **8**(1)**:** 59-71.
735

736   53.   Ozcan L, Ergin AS, Lu A, Chung J, Sarkar S, Nie D *et al.* Endoplasmic reticulum stress
737   plays a central role in development of leptin resistance. *Cell Metab* 2009; **9**(1)**:** 35-51.
738

739   54.   Campbell JN, Macosko EZ, Fenselau H, Pers TH, Lyubetskaya A, Tenen D *et al.* A
740   molecular census of arcuate hypothalamus and median eminence cell types. *Nat*
741   *Neurosci* 2017; **20**(3)**:** 484-496.
742

743   55.   Hirata T, Nakazawa M, Muraoka O, Nakayama R, Suda Y, Hibi M. Zinc-finger genes Fez
744   and Fez-like function in the establishment of diencephalon subdivisions. *Development*
745   2006; **133**(20)**:** 3993-4004.
746

747   56.   Kiyama T, Chen CK, Wang SW, Pan P, Ju Z, Wang J *et al.* Essential roles of
748   mitochondrial biogenesis regulator Nrf1 in retinal development and homeostasis. *Mol*
749   *Neurodegener* 2018; **13**(1)**:** 56.
750

751

752

753   **FIGURE LEGENDS:**

754   **Figure 1: An integrative functional genomics approach to model the differentiation of**

755   **hypothalamic neurons.** (A) Schematic of the study design. ESC, HP, and HNs were used to

756   generate RNA-seq, ATAC-seq, and Capture C profiles, which we compared to the GWAS

757   signals mined using our variant to gene mapping approach. (B) Expression level of NKX2-1

758   determined by RNA-seq. (C) Accessibility change of OCR located in the NKX2-1 promoter over

759   the course of HN differentiation. (D) A 600kb region around the NKX2-1 gene in ESCs (teal),

760   HPs (purple), and HNs (orange). The peaks track represents the ATAC-seq coverage, where

761   higher peaks depict increased accessibility (open chromatin), and arcs represent significant

762   (Chicago Score > 5) interacting regions between the NKX2-1 promoter. (E) Comparison of HN

763   expression profile to median GTeX database, scores give the spearman correlation coefficient

764   of the top 16,953 genes expressed in both datasets.

765
766   **Figure 2: Gene expression and chromatin architecture underlies during hypothalamic**

767   **neuron differentiation**: (A) Heatmap of the scaled expression values (z-score) of each RNA-

768   seq sample of ESC, HP, and HN for the 15,808 genes with differential expression in at least one

769   stage of differentiation. Red indicates higher relative expression, while blue indicates lower

770   relative expression. From hierarchical clustering, the genes were assigned to six groups based

771   on their expression pattern pointed to being either specifically enriched or depleted in one

772   condition (blue, orange, pink, green, yellow, gray bars). The log transformed condition-wise max

773   TPM value for each gene is plotted where white indicates lower expression and orange

774   indicates higher expression compared to other genes. (B) The global average of expression

775   values (TPM) the genes in each cluster. The central line of each boxplot represents the median,

776   with edges representing the 25 and 75 percentiles, and whiskers represent the 5 and 95

777   percentiles. (C) Sets of OCRs were assigned based on location: promoter OCRs (solid), PIR-

778   OCRs (hashed), and nonPIR-OCRs the remaining OCRs that were not annotated to a gene.

779   Bottom: The distribution of number OCRs annotated to each set per celltype. (D) Volcano plot

780   depicting the global genome-wide significant differentially accessible OCRs in the transition from

781   ESC to HP (left) and HP to HN (right). (H) Distribution of chromatin accessibility fold-change of

782   cisRE annotated to DE gene clusters (**Figure 2A**).

783
784   **Figure 3: TF footprinting and analysis: (A)** The summary of TF binding site prediction, the

785   Purity scores calculated by PIQ (Purple/Yellow), the number of TF binding sites passing purity

786   cutoff > 0.7 threshold (Grayscale), $\log_2$ transformed expression of the respective TF in HN

787   (white /Orange). (B) Grouping number of TF binding sites family. Darker color indicates a higher

788   number of predicted binding sites. (C) Enrichment of TF binding sites in putative cREs

789   compared to other OCRs in each cell type adjusted for GC content and read count.

790
791   **Figure 4: Putative cis regulatory elements are enriched for GWAS signals of complex**

792   **traits**. (A) Partitioned LDscore regression of ESC, HP, and HN cRE against the indicated

793   genome-wide signal. Circle size indicates the enrichment of estimated heritability and color

794   indicates statistical significance. (B) Comparison of genes implicated in our variant to gene

795   mapping analysis for each GWAS. Dots and lines indicate the intersect of the set of genes found

796   in each GWAS. Top: the number of genes in different overlapping sets. Right: the number of

797   SNPs detected in each GWAS. (C) Proportion of genes from our variant to gene mapping

798   located in each DE cluster (**Figure 2A**) or non-DE genes (black). (D) *FEZF1* genomic locus with

799   interactions connecting to a distal cRE. The SNP is located in a putative NRF1 motif. (F)

800   Genome track for the BDNF locus. Multiple proxies in open regions are shown. Two proxies

801   located in putative TF motifs for CUX1/2 and MAF::NFE2 are shown. Abbreviations: AAM: Age

802   at Menarche; BMI: Body Mass Index; LeptinBMIadj: Circulating Leptin levels adjusted for BMI;

803   MDD: Major Depressive Disorder; PTSD: Post Traumatic Stress Disorder; Sleep L5 Time; Least

804   active 5 hours; Sleep N Episodes = Number of Sleep episodes; Sleep Dur; Sleep Duration;

805   Sleep Dur SD: Sleep Duration Standard Deviation; Sleep Diurnal Inact: Sleep Diurnal Inactivity;

806   T2D: Type 2 Diabetes.

807

808

809   **SUPPLEMENTAL FIGURE LEGENDS:**

810

811   **Supplemental Figure 1:** (A) The first and second principal components of RNA-seq and (B)

812   ATAC-seq samples. The variance explained by each principal component is indicated on each

813   axis. (C) Pairwise Pearson correlation coefficients of RNA-seq and (D) ATAC-seq samples. (E)

814   The expression (TPM) of selected marker genes during the three stages of differentiation.

815

816   **Supplemental Figure 2:** (A) Counts of the number of genes assigned to each cluster. (B) The

817   relative number of genes annotated as either protein coding, lincRNA, or other annotation (e.g.

818   small noncoding RNAs, pseudogenes). (C) Top 10 GO term or (D) REACTOME pathway

819   enrichment of each gene group. The bar's length depicts the -log10 transformed FDR adjusted

820   p-values. The red dotted line indicates the threshold for statistical significance (FDR < 0.05).

821

822   **Supplemental Figure 3:** (A) The genomic location of all OCRs and the (C) PIR-OCRs. (B)

823   Distribution of the number of cRE at each stage. (C) Quantification of the distance (log10) of

824   each cRE (D) Distance between fragments detected by Promoter-focused Capture C at 1-

825   fragment and 4-fragment resolutions from each stage of differentiation. (E) Distribution of the

826   number of PIR-OCRs linked to each gene in each stage of differentiation. (F) The distribution of

827   the number of cRE binned by expression quantile. (G) Chromatin accessibility profile relative to

828   all transcription start sites. (H) Percentage of fragments with interactions to other promoter bait

829   fragments (bait-to-bait) at each resolution.

830

831   **Supplemental Figure 4:** (A) Comparison between ESC and HP enrichment scores (-

832    $\log_{10}$(BiFET P value)). (B) Comparison between HN and HP enrichment scores. (C-F)

833   Expression (log$_2$TPM) of top enriched TFs biased toward HPs compared to ESC (C), HP

834   compared to HN (D) and HN to HP (E), and TFs highly enriched in both HPs and HNs (F).

835

836

837

838 **SUPPLEMENTARY TABLES:**

839

840 **Supplemental Table 1:** Results from Spearman Correlation tests between HNs and GTeX

841 tissues.

842 **Supplemental Table 2:** Enriched TF motifs in cRE compared to unannotated OCRs

843 **Supplemental Table 3:** Results for partitioned LD score regression for the indicated genotypes.

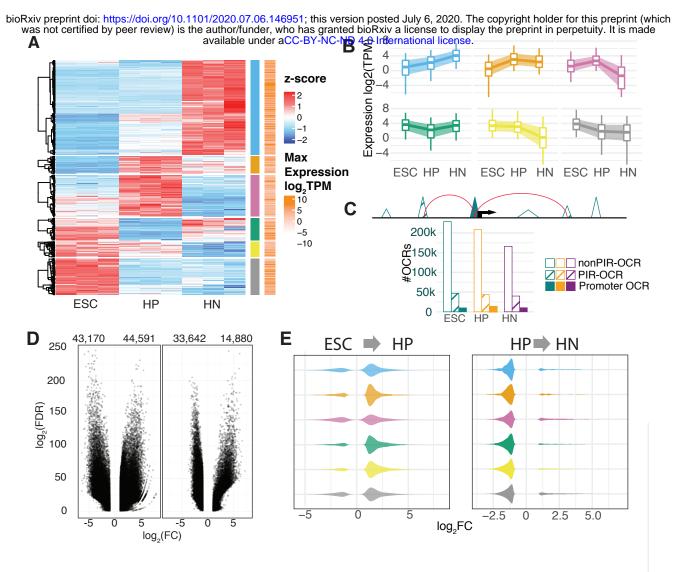844 **Supplemental Table 4:** DAVID annotations of implicated genes for each trait.

845 **Supplemental Table 5:** Gene annotation to each indicated proxy and sentinel SNP.

846 **Supplemental Table 6:** Overlap between GWAS traits, 0 = not implicated, 1 = implicated.
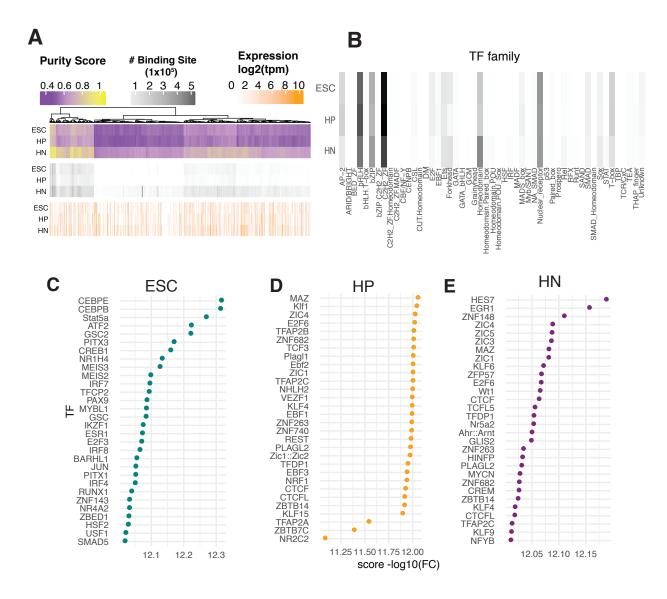
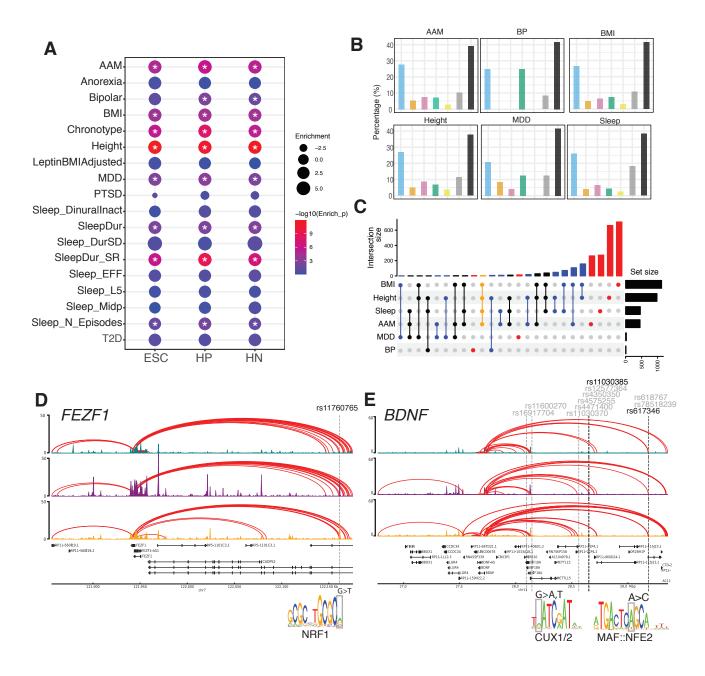847 **Supplemental Table 7:** Annotation of SNPs located in TF binding sites.

848 **Supplemental Table 8:** Enriched GO terms of genes implicated by our V2G pipeline mapping
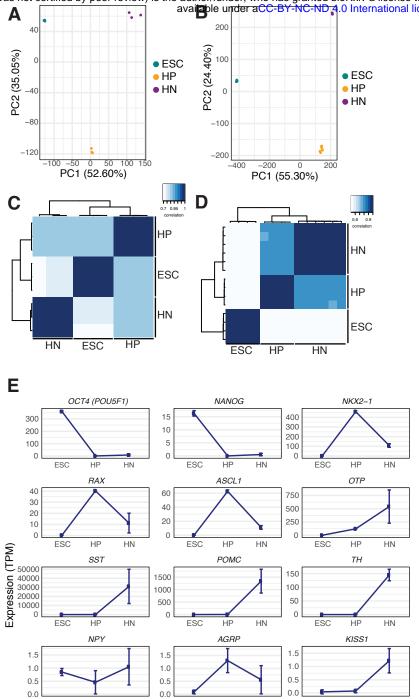
849 specifically in HPs and HNs.

A

Day 0 ESC

Day 12 HP

Day 27 HN

RNA-SEQ
CAPTURE-C
ATAC-SEQ

Differentiation / Function

GWAS

Causal gene annotation

B *NKX2–1*

C *NKX2-1 promoter*

D NKX2-1

E GTEX Correlation

7.5

Trait	GWAS Signals	In reference panel	Unique proxies	Open proxies with *cis* interactions[a]	Sentinels with *cis* interactions[a]	Contacted genes[b]	Unique contacted genes[b]
Height	3290	3254	165039	1081, 1098, 1073	594, 562, 564	543, 439, 507	981
BMI	941	928	63148	619, 474, 575	271, 211, 248	800, 526, 802	1126
AAM	499	463	13040	242, 194, 258	109, 105, 114	365, 267, 268	452
Sleep	459	445	26445	313, 223, 305	98, 84, 99	361, 212, 319	461
MDD	44	42	3059	22, 18, 17	11, 10, 8	15,16,16	24
BP	39	30	1697	1, 1, 1	1, 1, 1	2, 0, 11	12

[a]ESCs, HPs, and HNs, respectively

[b]Expressed at TPM > 1