Regulatory risk loci link disrupted androgen response to pathophysiology of Polycystic Ovary Syndrome

Jaya Srivastava and Ivan Ovcharenko*

Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD, 20892, USA.

* To whom correspondence should be addressed: ovcharen@nih.gov

Keywords: Polycystic Ovary Syndrome (PCOS); regulatory genomics; enhancer variants; deep learning, artificial intelligence; disease-causal noncoding variants

Abstract

A major challenge in deciphering the complex genetic landscape of Polycystic Ovary Syndrome (PCOS) lies in the limited understanding of how susceptibility loci drive molecular mechanisms across diverse phenotypes. To address this, we integrated molecular and epigenomic annotations from proposed causal cell-types and employed a deep learning (DL) framework to predict cell-type-specific regulatory effects of PCOS risk variants. Our analysis revealed that these variants affect key transcription factor (TF) binding sites, including NR4A1/2, NHLH2, FOXA1, and WT1, which regulate gonadotropin signaling, folliculogenesis, and steroidogenesis across brain and endocrine cell-types. The DL model, which showed strong concordance with reporter assay data, identified enhancer-disrupting activity in approximately 20% of risk variants. Notably, many of these variants disrupt TFs involved in androgen-mediated signaling, providing molecular insights into hyperandrogenemia in PCOS. Variants prioritized by the model were more pleiotropic and exerted stronger downregulatory effects on gene expression compared to other risk variants. Using the IRX3-FTO locus as a case study, we demonstrate how regulatory disruptions in tissues such as the fetal brain, pancreas, adipocytes, and endothelial cells may link obesity-associated mechanisms to PCOS pathogenesis via neuronal development, metabolic dysfunction, and impaired folliculogenesis. Collectively, our findings highlight the utility of integrating DL models with epigenomic

data to uncover disease-relevant variants, reveal cross-tissue regulatory effects, and refine mechanistic understanding of PCOS.

Keywords

Polycystic Ovary Syndrome (PCOS); regulatory genomics; enhancer variants; deep learning, artificial intelligence; disease-causal noncoding variants

38 **Introduction**

39 Polycystic Ovary Syndrome (PCOS) is a multifactorial endocrine disorder characterized by
40 abnormal LH:FSH (Luteinizing hormone: Follicle Stimulating hormone) ratios and
41 elevated androgen levels, leading to anovulation, polycystic ovaries, and various
42 hyperandrogenism-related comorbidities (1–3). The reproductive abnormalities in PCOS
43 stem from disruptions in the hypothalamic-pituitary-gonadal (HPG) axis, which also
44 contributes to other conditions like oligomenorrhoea, ovarian insufficiency, infertility,
45 hyper- and hypogonadism, and endometriosis (4). This overlap in clinical features
46 complicates PCOS diagnosis, prompting the establishment of multiple diagnostic criteria
47 by the NIH, Rotterdam, and the Androgen Excess and PCOS Society. Based on a
48 consensus, diagnosis relies on clear indications of hyperandrogenism, polycystic ovarian
49 morphology and ovulatory dysfunction (3,5).

50

51 Decades of research on molecular mechanisms underlying the disease have identified
52 impaired folliculogenesis and enhanced steroidogenesis in theca and granulosa cells
53 (GCs) as key contributors to PCOS development (6). These pathways are
54 spatiotemporally regulated by LH and FSH, secreted by the pituitary gland in response to
55 hypothalamic Gonadotropin Release Hormone (GnRH) based stimulation (7). Moreover,
56 PCOS often coincides with hyperinsulinemia, though the molecular origins of the
57 association between the two is still being investigated. Hyperinsulinemia worsens

58 hyperandrogenism by affecting adrenal androgen production and reducing sex hormone-
59 binding globulin (SHBG) levels in the liver (8) and may also contribute to the metabolic
60 co-morbidities such as obesity, type-II diabetes and liver dysfunction (3,9).
61 The genetic basis of PCOS is thought to involve impaired regulation of the HPG axis (10).
62 Polymorphisms in genes coding for kisspeptin (Kiss1, an upstream regulator of GnRH),
63 GnRH receptor, Anti-Mullerian Hormone (AMH), LH, FSH, and their receptors have been
64 linked to impaired signaling in PCOS patients (11,12). However, except for *AMHR*, no
65 functional studies directly connect these polymorphisms to the PCOS phenotype (13).
66 Twin studies suggest an estimated 80% heritability [13], highlighting the need for
67 detailed investigations into the molecular mechanisms underlying PCOS etiology (14).
68
69 PCOS genome-wide association studies (GWAS) across diverse populations have
70 identified novel disease loci, including plausible candidates such as *FSHR, FSHB*, and
71 *LHCGR*, in addition to several others with no direct association with disease phenotypes
72 (10). Variants in the loci *of THADA, DENND1A, IRF1, FTO* etc., have significant correlations
73 with the disease manifestation but have not been contextually studied. On the other
74 hand, polymorphisms in reproductive hormone receptors, including androgen (AR) and
75 estrogen receptors (*ESR1/2*), have been implicated in PCOS phenotypes (15,16);
76 however, these associations have not been identified in GWAS studies. These
77 observations align with the omnigenic model of complex trait regulation, where core

genes directly influence the phenotype, while peripheral genes contribute through cell-type-specific regulatory networks (17). Variants exert their net impact through complex genomic and epigenomic interactions, manifesting as GWAS association signals. This highlights the need for further investigation into the regulatory networks governing phenotypic complexity of PCOS. In this context, several isolated studies offer insights into the involvement of GWAS-identified genes in PCOS pathophysiology (10). For instance, studies show that *ERBB4* and *GATA4* regulate folliculogenesis (18,19), *ZNF217* regulates androgen production in theca cells (20), and *HMGA2* promotes granulosa cell proliferation (21). Interestingly, some genes exhibit pleiotropy depending on cell-type context; *HMGA2* also regulates adipogenesis (22), and FSH influences bone density and adipose mass (4). These findings suggest that PCOS-associated variants across multiple loci impact different cell-types through hitherto unexplored molecular mechanisms, contributing to phenotypic comorbidities.

In this study, we performed a functional assessment of PCOS susceptibility loci by integrating epigenomic data, functional assays, and a deep learning (DL)-based approach to identify causal single nucleotide variants (SNVs) across eleven disease-associated cell types. We further investigated their potential influence on the molecular mechanisms underlying PCOS etiology. This approach facilitated the identification of key transcription factors (TFs) involved in folliculogenesis, androgen-mediated signaling, and ovarian development, whose binding sites are predicted to be disrupted by causal variants. Using

98 the well-characterized regulatory locus of *IRX3*, we demonstrate how DL models
99 combined with prior knowledge of key PCOS TFs can effectively prioritize causal variants.

100 **Results**

101 **A majority of PCOS risk SNVs lie in regulatory regions and are enriched in**
102 **neuroendocrine cell-types**

103 To conduct a comprehensive analysis of the regulatory features of PCOS risk loci, we
104 identified 91 single nucleotide variants (SNVs) from twelve GWAS studies (Table S1). This
105 set was expanded to 1,472 SNVs, referred herein as pcosSNVs, by including variants in
106 linkage disequilibrium (LD) with GWAS-identified variants across all superpopulations
107 (African, American, South and East Asian, European) with an $r^2 \geq 0.8$, obtained from SNiPA
108 (23). Adjacent variants within 100kb were merged to define 50 genetic loci, named based
109 on the nearest gene and/or previously associated genes in the literature (Figure 1A, Table
110 S2). Most of these variants are located within intronic regions, with the highest density
111 observed in the *DENND1A* and *AOPEP* loci (Figure 1A). We then assigned target genes
112 using the ENCODE-rE2G model (https://github.com/EngreitzLab/ENCODE_rE2G) that
113 predicts enhancer-gene interactions across various cell types by integrating enhancer
114 activity, 3D chromatin interactions, and DNase I hypersensitivity maps. Using a threshold
115 of 0.8 for the rE2G score predicted from the logistic regression model, we obtained 97
116 target genes (Table S3), many of which were not previously linked to PCOS. These genes

117 are significantly enriched in pathways related to cell development, differentiation, and
118 apoptosis (hypergeometric p-value 10^{-6}, Figure 1B), highlighting their potential roles in
119 the biological mechanisms underlying the five developmental stages of folliculogenesis
120 in oocytes and granulosa cells (24).

121 The diverse phenotypic comorbidities associated with PCOS typically manifest as either
122 metabolic or reproductive abnormalities. To investigate potential differences between
123 these subtypes, we categorized pcosSNVs into metabolic or reproductive groups based
124 on two defined criteria: (1) trait descriptions provided in the GWAS summary statistics
125 (Table S1), and (2) alignment of phenotypic characteristics reported in the respective
126 GWAS with known subtype-specific features—namely, elevated AMH, LH, and SHBG
127 levels for reproductive, and high BMI, insulin, or glucose levels for metabolic—as
128 described in prior studies (25,26). Based on the sub-phenotype associations of GWAS
129 susceptibility variants, we classified fifteen loci as metabolic and sixteen as reproductive
130 (Table S4). To elucidate the regulatory mechanisms distinguishing the two subtypes, we
131 examined the differential enrichment of TFBSs in each subtype (Figure S1A, Methods).
132 The metabolic pcosSNVs showed significant enrichment for TFBSs of NR1D1 and *RXRB*,
133 which regulate adipocyte differentiation (27,28), and steroid hormone nuclear receptors
134 *RXRB*, and *THRA* (28) (binomial p-value < 0.01 vs reproductive pcosSNVs). Interestingly,
135 binding sites of TFs involved in the development and differentiation of GCs during
136 folliculogenesis, such as *CPEB1* and *FOXL2*, are enriched in the metabolic subtype. This

137 suggests their role in also driving metabolic abnormalities associated with PCOS, aligning
138 with previous observations where CPEB1 has also been shown to associate with obesity
139 (29). Conversely, the reproductive subtype is enriched for TFBSs of *NR5A1*, and *NR1H4*
140 (binomial p-value < 0.01 vs metabolic pcosSNVs), all of which have well-documented
141 roles in steroidogenesis and folliculogenesis (30,31). Furthermore, the two subtypes are
142 enriched for variants from distinct GWAS trait groups, correlating with their phenotypic
143 traits (Figure S1B, Methods). For example, reproductive subtype variants exhibit
144 significant enrichment for endocrine traits like endometriosis and uterine fibroids
145 (binomial p-value < 10^{-15} vs pcosSNVs), while metabolic subtype variants show significant
146 enrichment for obesity, BMI, and cholesterol (>5-fold, binomial p-value < 10^{-30} vs
147 pcosSNVs).

148 We next investigated the functional impact of pcosSNVs through their association with
149 changes in gene expression characterized by the GTEx consortium (32). Among 1,472
150 pcosSNVs, 832 overlapped with cis-eQTLs, termed eVariants (Table S5). Two-thirds of
151 these pcosSNVs are shared across multiple cell-types (we use the term cell-types herein
152 synonymously with tissues defined in GTEx), meaning they influence gene expression in
153 multiple cell types. In contrast, the remaining variants, such as those in the *FSHR* and
154 *ERBB4* loci, are cell-type specific (Figure S2a). The number of target genes scaled almost
155 linearly with the number of affected cell types (Spearman correlation: 0.71, Figure S2b),
156 suggesting that shared eQTLs may contribute to distinct cell-type specific regulatory

157 networks by regulating different genes in different cell-types. Notably, eVariants in the
158 locus of the GATA4 gene were linked to 39 genes across 49 cell types (Figure S2b, Table
159 S6). 4% of eVariants targeted long non-coding RNAs (Table S5), suggesting their role as
160 potential trans-eQTLs (33). Meanwhile, the remaining 640 pcosSNVs that were not
161 identified as GTEx eVariants belong to 18 susceptibility loci, including that of *CNTNAP5*,
162 *ASIC2*, and *CDH1*, (Table S6), likely due to low target gene expression or restricted
163 function in specific cell states or developmental stages in the dynamic transcription
164 landscape that not captured in bulk tissue analysis (32,34). These pcosSNVs may play key
165 spatiotemporal roles in mediating GnRH response in the hypothalamus, pituitary
166 regulation, and follicular phase progression in PCOS (32).

167 We also examined the enrichment of PCOS eVariants across GTEx cell types. Compared
168 to a randomly selected set of 832 eQTLs (excluding PCOS eVariants), PCOS eVariants were
169 found to be enriched in brain cell types, reproductive hormone-producing tissues such as
170 the ovary and adrenal gland, as well as hormonally influenced tissues like the breast and
171 prostate (binomial p-value < 0.001, Figure 1C). Since many PCOS eVariants are shared
172 across cell-types, they likely influence regulatory networks by mediating interactions
173 between cell-type-specific and ubiquitous transcription factors (TFs), thereby invoking
174 cell-type-specific regulatory pathways that contribute to distinct phenotypic outcomes
175 in different cellular contexts (35). For example, *PPARG*, a susceptibility locus, plays a
176 central role in regulating lipid metabolism, adipocyte differentiation, gluconeogenesis,

177 folliculogenesis, and steroidogenesis through multiple TFs that are critical regulators of
178 these biological processes (www.kegg.jp/pathway/map=map03320) (36,37). This
179 suggests that causal variants within this locus may contribute to distinct phenotypic
180 outcomes through pleiotropic effects across multiple cell types. Given these
181 complexities, investigating disease-causal variants and their cell-type-specific effects on
182 downstream signaling pathways may help elucidate the mechanisms underlying the
183 diverse phenotypic manifestations of PCOS.

184 **A deep learning model for prioritizing disease causal variants across causal cell-types**

185 The challenge of characterizing the cell-type-specific impact of thousands of
186 susceptibility variants in complex traits and diseases has led to the development of
187 computational approaches for inferring causal variants. DL models have been particularly
188 effective in predicting variant effects on gene regulation by integrating diverse cell-type-
189 specific epigenomic features (38,39). We previously developed a convolutional neural
190 network based DL model, TREDNet, which can predict the effects of non-coding variants
191 on enhancer activity (40). This two-phase DL model was demonstrated as a successful
192 approach in prioritizing causal variants of type 2 diabetes and autism (40,41). Building on
193 its success, we applied TREDNet to investigate the regulatory mechanisms underlying
194 PCOS.

195

196 We adapted TREDNet to predict allele-specific enhancer activity of PCOS-associated
197 SNVs (pcosSNVs) across causal cell types implicated in PCOS. Based on the biological
198 origins of folliculogenesis, androgenesis, and signaling pathways involving SHBG, insulin
199 signaling, and adipogenesis that are known to be disrupted PCOS, we identified eleven
200 causal cell types for analysis with available epigenomic profiles: KGN (used as a proxy for
201 granulosa cells due to limited epigenomic data from primary granulosa cells), ovary, fetal
202 brain, adrenal gland, pancreas, adipocytes, liver, brain microvascular endothelial cells
203 (BMEC), mammary epithelial cells, and human umbilical vein endothelial cells (HUVEC).
204 In the absence of epigenomic profiles from the pituitary and hypothalamus, we included
205 fetal brain, and selected BMEC, mammary epithelial cells, and HUVEC as granulosa cell
206 proxies based on their epithelial or endothelial characteristics and similar H3K27ac
207 profiles (Jaccard similarity index, Table S7) (42). Additionally, we incorporated WTC11, a
208 developmental cell line, to capture causal variants active during early development, as
209 fetal development has been implicated in PCOS onset later in life (43). We trained our
210 model using cell-type-specific putative enhancers identified by co-occurrence of the
211 H3K27ac histone modification with DNase hypersensitive sites (DHS), as proxies for
212 active enhancers (Table S8, Methods). The DL models demonstrated robust
213 performance, achieving an area under the receiver operating characteristic curve
214 (auROC) ranging from 0.9 to 0.98 and an area under the precision-recall curve (auPRC)
215 ranging from 0.54 to 0.84 across the eleven cell types (Figure 2A).

216 To evaluate TREDNet's ability to identify causal variants, we examined the correlation
217 between TREDNet-predicted differences in allele-specific enhancer activity and those
218 determined through a massively parallel reporter assay (MPRA) in the developing human
219 brain and stem cell-derived adipocytes using our model trained on fetal brain and
220 adipocytes (44,45). We compared allele-specific TREDNet scores across all assayed
221 alleles and those showing significant changes in reporter activity (Methods) and
222 observed a significantly higher fold change in TREDNet scores for the latter group (Mann-
223 Whitney test p-value = 0.001 for adipocytes and 10^{-5} for fetal brain, Figure 2B). These
224 findings highlight TREDNet's robustness in predicting causal variants across different cell
225 types.

226 Next, we evaluated the impact of pcosSNVs within active regulatory regions across the
227 eleven selected cell-types, scoring both reference and risk alleles in all cell-type-specific
228 models (Methods). For pcosSNVs located in active regulatory regions marked by
229 H3K4me1, H3K27ac, or DNase/ATAC-Seq, we classified strengthening alleles as those
230 with scores below the threshold (determined at 10% FDR) for the reference allele and
231 above for the risk allele, while damaging alleles followed the opposite criterion. Applying
232 this approach, we identified 309 pcosSNVs with predicted allelic differences in activity,
233 termed reSNVs (Table S9). These reSNVs were significantly enriched in conserved
234 elements compared to both pcosSNVs and 13 million common SNVs from the 1000
235 Genomes catalog (binomial p-value = 0.0002 and 10^{-9}, respectively, Figure 2C).

236 To assess the regulatory impact of reSNVs, we examined the enrichment of TFBSs. The
237 regulatory effect of a TF was quantified by comparing the density of its binding motifs
238 overlapping with reSNVs to a background set of control SNVs. Specifically, we quantified
239 the abundance of transcription factor binding sites (TFBSs) affected by these variants and
240 compared it to a control set of TFBSs corresponding to 71,000 SNVs located within 100
241 kb of pcosSNVs (Methods). This localized background enabled us to investigate the
242 regulation of target genes within the context of PCOS-specific biological processes,
243 particularly for ubiquitously expressed genes. Several TFs showed significant enrichment
244 at reSNV loci, including *FOXA1*, a pioneer factor in estrogen and androgen signaling (46);
245 *LHX4*, involved in pituitary development (47); *NHLH2*, associated with GnRH signaling
246 (48); *WT1*, a regulator of granulosa cell proliferation (49); PLAG1, involved in oocyte
247 reserve maintenance (50); and *NR4A1*, which regulates steroidogenesis (51)
248 (hypergeometric p-value < 10^{-2}, Figure 2D). Notably, we observed a 2.6 fold enrichment
249 of PPARG binding sites, a significant finding given PPARG's role as a known susceptibility
250 locus for PCOS. We also found enrichment of TFs associated with neuronal signaling, such
251 as TBX21, POU6F1, and NKX6.2. While not previously linked to PCOS, these TFs represent
252 promising candidates for involvement in neuroendocrine regulation. These findings
253 highlight the capacity of our model to identify transcriptional regulators with potential
254 functional roles in the diverse phenotypic manifestations of PCOS.

255 Our deep learning-based approach identified 20% of the pcosSNVs as potential
256 regulatory variants, effectively narrowing down causal variants in loci such as *ERBB4*,
257 *LHCGR, MC4R*, etc (Figure S3). For example, among 51 variants in the *ERBB4* locus, we
258 identified rs79230362 as an enhancer-disrupting variant in HUVEC cells (Table S9). This
259 variant is in LD with the GWAS SNP rs113168128 and is predicted to disrupt the binding
260 site of the *ELK1:SREBF2* motif complex (Figure S4). Given the established role of *SREBF2*
261 in steroidogenesis (52) and the highly cell-type-specific expression of ELK1 in granulosa
262 cells (Figure S4), this variant likely affects *ERBB4* expression, a key regulator of the oocyte
263 microenvironment during folliculogenesis (18). Similarly, in the *MC4R* locus, we
264 identified rs17773430 as a causal enhancer-disrupting variant in WTC11 cells (Table S9).
265 *MC4R* is a critical component of the melanocortin pathway and a well-established obesity
266 susceptibility gene that is also linked to PCOS. Knockout studies of *MC4R* in mice result
267 in both obesity and infertility phenotypes, highlighting shared regulatory architectures
268 underlying these conditions (53). rs17773430 is predicted to disrupt the binding site of
269 *TBX2/TBXT*, TFs responsible for the development of hypothalamus-pituitary axis (30).
270 Given that reduced *MC4R* levels are associated with lower LH levels (54), this variant
271 likely contributes to PCOS etiology through its impact on HPG axis.
272 On the other hand, multiple reSNVs were identified in the locus of *DENND1A*, *FTO* and
273 *MAPRE1* (Figure S5). The significant overlap of reSNVs in *DENND1A* and *MAPRE1* locus
274 with active regulatory regions in the fetal brain and WTC11 suggests their potential role

275 in disease manifestation during early development. Notably, a reSNV in the *MAPRE1*
276 locus, rs187178, was validated as an enhancer-disrupting variant in the fetal brain and
277 functions as an eQTL for the neighboring gene *DNMT3B*, which regulates dynamic
278 methylation transitions during folliculogenesis (24). In total, we identified 12 reSNVs that
279 have been experimentally validated as enhancer disrupting variants in adipocytes and
280 fetal brain through MPRA studies (Table S10) (44,45).
281 Of note, epigenomic data from fetal brain used by the DL model failed to capture the
282 regulatory impact of pathogenic variants in the *FSHB* locus, including rs10835638 and
283 rs11031006, which have been experimentally shown to reduce *FSHB* expression
284 restricted to the pituitary gland (55). This underscores the necessity of incorporating
285 additional, relevant cell types for a more comprehensive study of the regulatory
286 landscape of PCOS, when experimental characterization of chromatin marks becomes
287 available for these cell types.

288 **reSNVs are more likely to exert pleiotropic effects across multiple cell types by**
289 **downregulating the expression of their target genes**

290 We further explored the functional impact of reSNVs by examining their association with
291 gene expression using eQTL data from GTEx. Among reSNVs that also act as eVariants,
292 hereafter referred to as reVariants, we observed significantly greater enrichment in the
293 brain, liver, adrenal gland, and pancreas compared to pcosSNVs (Figure 3A), implicating

294 these tissues as key cell types affected by reSNVs. The proportion of causal cell types
295 impacted by reVariants was significantly higher than that impacted by otherSNVs (i.e.,
296 SNVs not prioritized by TREDNet in causal cell types) (Figure 3B, Mann-Whitney p =
297 5.78×10^{-3}). Within these enriched cell types, reVariants were associated with
298 significantly stronger downregulation of gene expression relative to otherSNVs, as
299 measured by normalized effect sizes from GTEx (Figure 3C, Mann-Whitney p =
300 1.88×10^{-9}). In contrast, no significant difference was observed in gene upregulation
301 effects (Figure 3C, Mann-Whitney p = 0.91). These findings suggest that reSNVs primarily
302 exert their regulatory effects through downregulation of gene expression.

303 The most significant downregulatory effect was observed at the RAB5B–SUOX–RPS26
304 locus, where reSNVs were linked to reduced expression of RPS26 in multiple cell-types
305 including the ovary, hypothalamus, and liver. RPS26 is a ubiquitously expressed
306 ribosomal protein whose downregulation in the ovaries impairs oocyte growth and
307 premature ovarian failure (56), a hallmark of PCOS. Notably, one reVariant in this region,
308 rs3741499, which shows a large negative effect size on RPS26 expression (Figure S6), is
309 predicted to disrupt binding of PROX1 (Table S9), a PCOS risk gene involved in lymphatic
310 vessel formation around oocytes (57), suggesting a plausible mechanism for impaired
311 oocyte maturation.

312 We further investigated pleiotropy by examining the ZBTB16 locus. Although no eQTLs
313 overlap with variants in this locus, they were predicted to exert strong differential
314 enhancer activity across multiple cell types (Table S9). Notably, rs1784692, located in an
315 intron of ZBTB16, demonstrated the highest predicted enhancer-strengthening effect in
316 the pancreas, adipocytes, WTC11, and liver (Figure 3D, Table S9). The T→C polymorphism
317 enhances *AR* receptor binding, suggesting a possible association of this locus with cell-
318 type-specific androgen response functions, such as insulin secretion in the pancreas (58),
319 and regulation of adipocyte differentiation (59). While ZBTB16 has not been previously
320 implicated in PCOS, its protein interaction network is enriched for components of
321 androgen signaling (Figure 3E). These observations suggest that ZBTB16 may act as a
322 susceptibility locus involved in androgen-mediated regulatory pathways disrupted in
323 PCOS.

324 In conclusion, reSNVs prioritized by TREDNet offer valuable insights into disease-
325 associated regulatory mechanisms and highlight the potential role of risk genes hitherto
326 uncharacterized in PCOS etiology.

327 **The FTO locus demonstrates disruption of an androgen mediated network pleiotropy**

328 The regulatory locus within the intronic region of *FTO* is a well-known susceptibility locus
329 with significant implications in obesity and diabetes. Notably, it has been experimentally
330 validated to function as a distal enhancer of *IRX3*, a TF in PCOS-associated susceptibility

331 loci (60,61). We hypothesized that this locus may have broader pleiotropic effects across
332 different cell types due to variations in the expression of *IRX3*, which may influence
333 multiple biological pathways (62). Interestingly, the PCOS susceptibility variants localize
334 in the genomic region regulating *IRX3* (chr16:53731249–54975288) (63), suggesting that
335 *IRX3* is likely the target gene of the PCOS susceptibility locus as well (Figure S7).

336 We identified 12 reSNVs exhibiting significant fold changes across nine cell types (Figure
337 S8). Among these, three variants— rs1421085, rs9940646 and rs9940128—have been
338 validated by MPRA studies to show allelic changes in enhancer activity in mouse
339 preadipocyte and neuronal cell lines (61), further supporting the predictive accuracy of
340 TREDNet in identifying causal variants. Interestingly, we predicted that rs1421085
341 additionally upregulates enhancer activity in BMEC by potentially modulating the binding
342 site of *ONECUT2* (Figure 4A), a suppressor of androgen receptor signaling which was
343 recently identified as a marker of follicle growth (64,65).

344 Additionally, we identified another variant within the same locus, rs8050136, which is
345 predicted as a causal variant in the pancreas and liver (Figure 4A, S8). This variant
346 functions as an eQTL for *IRX3* in the pancreas, where IRX3 regulates the conversion of
347 beta to epsilon cells, directly linking it to type 2 diabetes (66). Notably, rs8050136 is also
348 predicted to disrupt the binding site for *ONECUT1*, a transcription factor critical for
349 pancreatic development (Figure 4A). Together, these findings suggest that rs8050136

350 may serve as another causal variant for type 2 diabetes, possibly preferentially in PCOS
351 patients.

352 To address the association of these variants with PCOS, we focused on a previous study
353 that identified *IRX3* and another gene in this susceptibility locus, *IRX5*, as key regulators
354 of folliculogenesis in granulosa cells (67). Using evidence from granulosa like cells, BMEC
355 and HUVEC, we hypothesize that variants in this locus lead to impaired folliculogenesis,
356 consequently disrupting androgen production in the causal cell type—likely granulosa
357 cells—through the dysregulated action of *IRX3/IRX5*. This disruption in androgen
358 production may have pleiotropic effects on other cell types where these genes function
359 within the androgen-responsive network. In this regard, rs9940128 emerges as a
360 plausible causal variant as it forms chromatin contacts with promoters of *IRX3* and *IRX5*
361 (Figure 4B) and is predicted to cause a significant fold change in enhancer activity in
362 BMEC and HUVEC (Figure S8). Furthermore, the allelic effects of variants in this locus may
363 also impact *IRX3/5*-mediated functions in hypothalamic neurons (Figure S7), as
364 demonstrated in mice (61). To explore this further, we analyzed the impact of these
365 variants in fetal brain and found that rs3751812 is located within binding sites of T-box
366 family TFs (Figure 4A). Members of the T-box family play a critical role in the commitment
367 of hypothalamus and pituitary lineages from neuronal precursors (30,68). However,
368 given the short temporal window of expression of these TFs in neuronal development,
369 inferring causal mechanisms remains challenging. This highlights the necessity of using

370 epigenomic datasets across different developmental timepoints for a comprehensive
371 investigation.

372 **Discussion**

373 Our limited understanding of the regulatory landscape of PCOS stems from its complex
374 genetic architecture, which presents with heterogeneous phenotypes across different
375 cell types, individuals, and populations. This complexity has necessitated evolving
376 diagnostic criteria as our knowledge of the underlying pathophysiology expands. Several
377 key questions remain unresolved, including the genetic and molecular origins of
378 reproductive and metabolic dysfunction, the role of androgens and other hormones in
379 regulatory pathways, and the inheritance patterns affecting both males and females. To
380 date, GWAS have identified 50 genomic loci associated with PCOS across diverse
381 populations (Table S2). While the functional significance of genes such as *ERBB4*, *PPARG*,
382 and *IRX3* has been well established, leading to the use of their agonists as potential
383 treatments (18,37,67), the precise molecular mechanisms remain elusive. Additionally,
384 advancements in whole-genome and exome sequencing continue to uncover novel loci,
385 further complicating our understanding of PCOS and highlighting the need for a deeper
386 exploration of the core regulatory mechanisms driving its pathophysiology.

387 Leveraging extensive genetic and epigenetic data, we sought to identify key mechanisms
388 linking PCOS susceptibility loci to disease etiology. We found that reSNVs prioritized by

389 our model are significantly enriched for TFBSs associated with folliculogenesis, including
390 those of WT1, NHLH2, and FOXA1. Notably, reSNVs also show enrichment for the binding
391 sites of PROX1 and PPARG, both of which are also PCOS risk genes. These findings
392 underscore the importance of dissecting the underlying gene regulatory networks,
393 where disruptions at specific nodes (genes) or edges (regulatory interactions) may give
394 rise to a spectrum of molecular outcomes that contribute to the heterogeneity of PCOS
395 severity and phenotypic presentation. Our results also highlight the need for further
396 characterization of TFs, especially those involved in neuronal signaling, such as *TBX21,*
397 *LHX4*, etc., along with their interactions with hormonal receptors, to gain deeper insights
398 into cis- and trans- regulatory mechanisms disrupted in PCOS pathophysiology.

399 The established role of the HPG axis (69) in regulating circulating reproductive hormone
400 levels highlights the hypothalamus, pituitary, adrenal gland, and ovarian granulosa and
401 theca cells as key mediators of PCOS pathophysiology. However, PCOS manifestations
402 extend beyond the neuroendocrine system, impacting peripheral tissues such as the
403 pancreas, adipocytes, liver, and heart. This suggests that dysregulation of hormonal
404 signaling, particularly androgens, may have widespread effects through both direct and
405 pleiotropic mechanisms. Given the broad expression of the androgen receptor,
406 disruptions in androgen signaling may contribute to metabolic dysfunctions—such as
407 insulin resistance and altered adipogenesis—independent of classical reproductive
408 symptoms like oligomenorrhea. Our findings support this expanded framework and

409 reveal potential mechanisms by which altered androgen signaling leads to systemic
410 effects. Accordingly, we propose two categories of pathogenic cell types: (a) primary cell
411 types, involved directly in steroidogenesis, folliculogenesis, and reproductive hormone
412 biosynthesis; and (b) secondary cell types, which are affected by the pleiotropic activity
413 of risk variants or by downstream hormonal dysregulation (Figure S9). By prioritizing
414 variants that disrupt PCOS relevant TF binding sites at susceptibility loci, we highlight the
415 importance of TFs interacting with hormone receptors—particularly androgens—as key
416 modulators of PCOS-related dysfunction.

417 The identification of multiple reSNVs at several susceptibility loci is suggestive of
418 regulatory mechanisms wherein one-gene can be regulated by multiple enhancers,
419 according to which, the expression of a target gene can be influenced by more than one
420 variant (61,70). For example, two distinct variants in the *FSHB* locus, rs10835638 and
421 rs11031006, alter *FSHB* expression, ultimately contributing to infertility (55). These
422 variants may occur in different individuals, leading to distinct, individual-specific
423 phenotypes depending on the cell-type-specific networks they modulate in a pleiotropic
424 manner. In addition, the potential pleiotropic impact of disease-associated variants in
425 non-pathogenic cell types is often buffered by robust regulatory networks, preventing
426 overt disease manifestation. This suggests that assessing polygenic risk scores may be
427 necessary to fully understand their contribution to disease susceptibility. Given that
428 variants in the *FTO* locus have high minor allele frequencies (>0.4), which far exceed the

429 prevalence of PCOS, it is evident that the disease phenotypes emerge from the
430 cumulative effects of multiple dysregulated genes and pathways. Further investigations
431 into polygenic interactions and gene-environment influences will be essential to expand
432 our understanding of the complexity of PCOS.

433 The susceptibility loci of PCOS implicate genes such as *ZBTB16*, *AOPEP*, *THADA*, and
434 *CCDC91* (Figure 1A), which are ubiquitously expressed, raising the question of how
435 disease-specific variants selectively affect certain cell types. At the molecular level,
436 follicle progression involves signaling pathways like TGFβ, Hippo, Wnt, and mTOR, which
437 regulate fundamental processes such as cell proliferation, differentiation, and apoptosis
438 (7). Why, then, do complex diseases manifest in only a subset of susceptible cell types?
439 In the case of *ZBTB16*, we predicted that rs1784692 strengthens enhancer activity by
440 increasing the binding affinity of *AR*, thereby implicating *ZBTB16* in downstream
441 pathways of androgen signaling. This suggests that perturbations in disease-relevant TF
442 interactions, specific to causal cell types, disrupt molecular networks in a way that
443 surpasses compensatory mechanisms in other cell types, thereby making certain cells
444 uniquely vulnerable. Consequently, transcription factors act as primary responders to
445 disease-associated alterations, preceding the genes they regulate, and may therefore
446 serve as more informative markers of disease susceptibility than the genes themselves.

447 Our analysis of the PCOS regulatory landscape reveals unifying molecular mechanisms
448 underlying disease phenotypes. However, a more comprehensive understanding of gene
449 regulatory networks requires integrating epigenomic datasets from key pathogenic cell
450 types—such as the pituitary gland, granulosa, and theca cells, and potentially, the
451 hypothalamus—across follicular phases to map the spatiotemporal regulation of genes
452 involved in steroidogenesis and folliculogenesis. Despite the hypothalamus's central role
453 in the HPG axis, regulatory networks mediated by GnRH signaling remain poorly
454 understood. Disruptions in this pathway may explain the involvement of risk genes such
455 as *CNTNAP5*, *ASIC2*, and *CUX2*, potentially linking PCOS to prevalent mental health
456 disorders (3). Incorporating these datasets can enable the development of more inclusive
457 deep-learning models capable of predicting regulatory activity changes beyond enhancer
458 disruptions, offering deeper insights into PCOS pathophysiology.

459 Additionally, our PCOS subtype classification remains incomplete due to lack of data,
460 leaving some loci unassigned, which may exclude crucial transcription factors and
461 interactions essential for understanding regulatory networks. Lastly, our analysis of
462 causal variants was limited to those occurring within putative enhancers. However,
463 variants can impact gene regulation beyond enhancer activity. Variants located in
464 silencers or insulators may disrupt distal enhancer interactions, as observed with IRX3,
465 emphasizing the need for Hi-C data from pathogenic and affected cell types to resolve
466 target genes not identifiable through eQTL analysis. Lastly, a comprehensive approach

467 should also consider the trans-regulatory effects of risk variants—whether through TFs
468 encoded by susceptibility loci (*PROX1*, *SOX5/8*, *IRF1*) or non-coding RNAs that contribute
469 to epigenomic regulation of gene expression.

470 **Conclusions**

471 Our results provide valuable insights into molecular mechanisms underlying PCOS
472 etiology. Future *in vitro* and *in vivo* characterization will be essential to validate these
473 predictions, potentially paving the way for novel, symptom-targeted therapies for PCOS
474 patients.

475 **Methods**

476 **PCOS susceptibility loci**

477 PCOS GWAS summary statistics were obtained from the NHGRI-GWAS catalog. Variants
478 in LD were expanded and clustered into 50 loci based on 100kb proximity. Subtypes
479 identified for 38 GWAS variants (Table S4) were also assigned to their LD variants. The
480 risk allele from GWAS summary statistic served as the alternate allele for GWAS variants,
481 while the minor allele from the 1000Genomes catalog was assumed as risk allele for LD
482 variants. All analyses were conducted using the coordinates and datasets of GRCh38
483 reference genome.

484 **Transcription factor binding sites**

485 Transcription factor binding site (TFBS) regions were defined by extending variant sites
486 by 30 bp on each side. TF binding profiles from HOCOMOCO (71) and JASPAR non
487 redundant collection (72) were analyzed using FIMO with default parameters (73). Aside
488 from gain and loss of motifs, changes in motif scores were used to assess affinity
489 differences between reference and alternate alleles. A list of all the TFBSs gained, lost
490 and modulated for SNPs exhibiting significant fold change is provided in Table S9.

491 **Cell type specific DL models**

492 We used a two phase TREDNet model developed in our lab for cell-type specific enhancer
493 prediction (74). The first phase of the model was pre-trained on 4560 genomic and
494 epigenomic profiles, which included DHS, ATAC-Seq, Histone ChIP-Seq and and TF ChIP-
495 Seq peaks from ENCODE v4 (75). The second phase was fine-tuned to predict cell type
496 specific enhancers using training datasets described below. Chromosomes 8 and 9 were
497 held out for testing, chromosome 6 was used for validation and other autosomal
498 chromosomes were used to build the second phase model. The area under the ROC and
499 PRC curve for each of these models is provided in Figure 2A. The pre-trained phase-one
500 model has been deposited at https://doi.org/10.5281/zenodo.8161621.

501 Open chromatin (DHS or ATAC-Seq) and H3K27ac profiles for the causal cell-types were
502 downloaded from ENCODE (75) (Table S8). Positive datasets were defined as 2 kb regions
503 centered on DHS or ATAC-Seq peaks overlapping with H3K27ac (or H3K4me1 in fetal

504 brain) peaks of each cell type, excluding coding sequences, promoter proximal regions (
505 <2kb from TSS) and ENCODE blacklisted regions (76). A 10-fold control dataset was
506 generated for each cell-type using randomly sampled 2kb fragments of the genome,
507 excluding the positive dataset of that cell type and blacklisted regions.

508 Each 2 kb fragment received an enhancer probability score. Active enhancers were
509 predicted at a 10% FPR with a 1:10 positive-to-control ratio. Variant effects were assessed
510 by scoring 2 kb regions centered on each variant for reference and alternate alleles. A
511 significant enhancer activity change was defined as an alternate/reference score ratio
512 >1.2 or <0.8.

513 **Enrichment analysis of TFBSs**

514 We used command line FIMO (77) to scan vertebrate TF motifs from JASPAR (78) and
515 HOCOMOCO (79) databases along the sequences, applying a p-value threshold of 10^{-5}.
516 To identify TFs enriched in the loci of pcosSNVs, we generated a background set of SNVs
517 by extracting all variants from the 1000 Genomes Project within a 50 kb flanking region
518 of each pcosSNV. After excluding the pcosSNVs themselves and removing duplicates, this
519 resulted in a non-redundant background set of approximately 71,000 SNVs. Differential
520 enrichment of TFBSs between the metabolic and reproductive subtypes was assessed
521 using a binomial test, with normalized counts of a TF overlapping variants of one subtype

522 analyzed against the normalized counts of the same TF overlapping variants of the other
523 subtype as the background.

524 **GWAS trait enrichment**

525 Summary statistics for 25,649 traits were downloaded from the NHGRI-GWAS catalog.
526 Linkage disequilibrium (LD) variants for each GWAS SNP were identified using PLINK
527 (v1.9(80)) with an r^2 threshold of ≥ 0.8. Traits with at least 1,000 combined GWAS and LD
528 variants were retained for downstream enrichment analysis in reproductive and
529 metabolic SNV categories.

530 **Data and tools**

531 The H3K27ac peaks for KGN cells and adipocytes were sourced from literature (81,82).
532 The KGN wig file was converted to NarrowPeak format using UCSC BigWig tools (83) and
533 MACS peak calling software (84).
534 Motif logos were retrieved from HOCOMOCO database (79). Ontology enrichment of
535 pcosSNVs was performed using the Molecular Signatures Database (85). Protein
536 interaction networks and enriched pathways (Figure 3E) were obtained from STRING
537 database (86).
538 Evolutionary conservation of genomic regions was measured by their extent of overlap
539 with phastCons elements conserved across 30 primates

(https://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/ phastConsElements30way.txt.gz).

Acknowledgements

This work utilized the computational resources of the NIH HPC Biowulf cluster.

Data availability

Please see the section "Data and tools" and supplementary tables. The deep learning models for eleven cell-types trained in the study are deposited at https://doi.org/10.5281/zenodo.15041688.

Funding

This research was supported by the Division of Intramural Research of the National Library of Medicine (NLM), National Institutes of Health.

Authors' contributions

J.S. performed the computational analysis, analyzed the data, and prepared figures and tables. I.O. supervised the study. J.S. and I.O. wrote the manuscript.

555 ## References

556 1. Joham AE, Norman RJ, Stener-Victorin E, Legro RS, Franks S, Moran LJ, et al. Polycystic ovary
557 syndrome. Lancet Diabetes Endocrinol. 2022 Sep 1;10(9):668–80.

558 2. Palomba S, Piltonen TT, Giudice LC. Endometrial function in women with polycystic ovary
559 syndrome: a comprehensive review. Hum Reprod Update. 2021 Apr 21;27(3):584–618.

560 3. Stener-Victorin E, Teede H, Norman RJ, Legro R, Goodarzi MO, Dokras A, et al. Polycystic ovary
561 syndrome. Nat Rev Dis Primer. 2024 Apr 18;10(1):1–23.

562 4. Zaidi M, Yuen T, Kim SM. Pituitary crosstalk with bone, adipose tissue and brain. Nat Rev
563 Endocrinol. 2023 Dec;19(12):708–21.

564 5. McCartney CR, Marshall JC. Polycystic Ovary Syndrome. N Engl J Med. 2016 Jul 7;375(1):54–64.

565 6. Abedel-Majed MA, Romereim SM, Davis JS, Cupp AS. Perturbations in Lineage Specification of
566 Granulosa and Theca Cells May Alter Corpus Luteum Formation and Function. Front Endocrinol
567 [Internet]. 2019 [cited 2023 Dec 15];10. Available from:
568 https://www.frontiersin.org/articles/10.3389/fendo.2019.00832

569 7. Wang K, Li Y. Signaling pathways and targeted therapeutic strategies for polycystic ovary syndrome.
570 Front Endocrinol. 2023;14:1191759.

571 8. Liao B, Qiao J, Pang Y. Central Regulation of PCOS: Abnormal Neuronal-Reproductive-Metabolic
572 Circuits in PCOS Pathophysiology. Front Endocrinol. 2021;12:667422.

573 9. Zhang CH, Liu XY, Wang J. Essential Role of Granulosa Cell Glucose and Lipid Metabolism on
574 Oocytes and the Potential Metabolic Imbalance in Polycystic Ovary Syndrome. Int J Mol Sci. 2023
575 Nov 13;24(22):16247.

576 10. Dapas M, Dunaif A. Deconstructing a Syndrome: Genomic Insights Into PCOS Causal Mechanisms
577 and Classification. Endocr Rev. 2022 Nov 25;43(6):927–65.

578 11. Gorsic LK, Dapas M, Legro RS, Hayes MG, Urbanek M. Functional Genetic Variation in the Anti-
579 Müllerian Hormone Pathway in Women With Polycystic Ovary Syndrome. J Clin Endocrinol Metab.
580 2019 Feb 20;104(7):2855–74.

581 12. Caburet S, Fruchter RB, Legois B, Fellous M, Shalev S, Veitia RA. A homozygous mutation of GNRHR
582 in a familial case diagnosed with polycystic ovary syndrome. Eur J Endocrinol. 2017
583 May;176(5):K9–14.

584 13. Chen Y, Wang G, Chen J, Wang C, Dong X, Chang HM, et al. Genetic and Epigenetic Landscape for
585 Drug Development in Polycystic Ovary Syndrome. Endocr Rev. 2024 Jul 12;45(4):437–59.

586 14. Vink JM, Sadrzadeh S, Lambalk CB, Boomsma DI. Heritability of polycystic ovary syndrome in a
587 Dutch twin-family study. J Clin Endocrinol Metab. 2006 Jun;91(6):2100–4.

588 15. Wang F, Pan J, Liu Y, Meng Q, Lv P, Qu F, et al. Alternative splicing of the androgen receptor in
589 polycystic ovary syndrome. Proc Natl Acad Sci U S A. 2015 Apr 14;112(15):4743–8.

590 16. Muccee F, Ashraf NM, Razak S, Afsar T, Hussain N, Husain FM, et al. Exploring the association of
591 ESR1 and ESR2 gene SNPs with polycystic ovary syndrome in human females: a comprehensive
592 association study. J Ovarian Res. 2024 Jan 29;17(1):27.

593 17. Mathieson I. The omnigenic model and polygenic prediction of complex traits. Am J Hum Genet.
594 2021 Sep 2;108(9):1558–63.

595 18. Veikkolainen V, Ali N, Doroszko M, Kiviniemi A, Miinalainen I, Ohlsson C, et al. Erbb4 regulates the
596 oocyte microenvironment during folliculogenesis. Hum Mol Genet. 2020 Oct 10;29(17):2813–30.

597 19. LaVoie HA. The GATA-keepers of ovarian development and folliculogenesis. Biol Reprod. 2014
598 Aug;91(2):38.

599 20. Waterbury JS, Teves ME, Gaynor A, Han AX, Mavodza G, Newell J, et al. The PCOS GWAS Candidate
600 Gene ZNF217 Influences Theca Cell Expression of DENND1A.V2, CYP17A1, and Androgen
601 Production. J Endocr Soc. 2022 Jul 1;6(7):bvac078.

602 21. Li M, Zhao H, Zhao SG, Wei DM, Zhao YR, Huang T, et al. The HMGA2-IMP2 Pathway Promotes
603 Granulosa Cell Proliferation in Polycystic Ovary Syndrome. J Clin Endocrinol Metab. 2019 Apr
604 1;104(4):1049–59.

605 22. Xi Y, Shen W, Ma L, Zhao M, Zheng J, Bu S, et al. HMGA2 promotes adipogenesis by activating
606 C/EBPβ-mediated expression of PPARγ. Biochem Biophys Res Commun. 2016 Apr 15;472(4):617–
607 23.

608 23. Arnold M, Raffler J, Pfeufer A, Suhre K, Kastenmüller G. SNiPA: an interactive, genetic variant-
609 centered annotation browser. Bioinforma Oxf Engl. 2015 Apr 15;31(8):1334–6.

610 24. Zhang Y, Yan Z, Qin Q, Nisenblat V, Chang HM, Yu Y, et al. Transcriptome Landscape of Human
611 Folliculogenesis Reveals Oocyte and Granulosa Cell Interactions. Mol Cell. 2018 Dec 20;72(6):1021-
612 1034.e4.

613 25. Dapas M, Lin FTJ, Nadkarni GN, Sisk R, Legro RS, Urbanek M, et al. Distinct subtypes of polycystic
614 ovary syndrome with novel genetic associations: An unsupervised, phenotypic clustering analysis.
615 PLoS Med. 2020 Jun 23;17(6):e1003132.

616 26. van der Ham K, Moolhuijsen LME, Brewer K, Sisk R, Dunaif A, Laven JSE, et al. Clustering Identifies
617 Subtypes With Different Phenotypic Characteristics in Women With Polycystic Ovary Syndrome.
618 [cited 2025 Apr 15]; Available from: https://dx.doi.org/10.1210/clinem/dgae298

619 27. Hunter AL, Pelekanou CE, Barron NJ, Northeast RC, Grudzien M, Adamson AD, et al. Adipocyte
620 NR1D1 dictates adipose tissue expansion during obesity. Tontonoz P, James DE, editors. eLife. 2021
621 Aug 5;10:e63324.

622 28. Evans RM, Mangelsdorf DJ. Nuclear Receptors, RXR, and the Big Bang. Cell. 2014 Mar
623 27;157(1):255–66.

624 29. Li D, Chen J, Yun C, Li X, Huang Z. MiR-122–5p regulates the pathogenesis of childhood obesity by
625 targeting CPEB1. Obes Res Clin Pract. 2022 May 1;16(3):206–13.

626 30. Pontecorvi M, Goding CR, Richardson WD, Kessaris N. Expression of Tbx2 and Tbx3 in the
627 developing hypothalamic-pituitary axis. Gene Expr Patterns GEP. 2008 Jul;8(6):411–7.

628 31. Candelaria NR, Richards JS. Targeted deletion of NR2F2 and VCAM1 in theca cells impacts ovarian
629 follicular development: insights into polycystic ovary syndrome?†. Biol Reprod. 2024 Apr
630 1;110(4):782–97.

631 32. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working
632 Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH
633 Common Fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. Nature.
634 2017 Oct 11;550(7675):204–13.

635 33. Bai Y, Dai X, Harrison AP, Chen M. RNA regulatory networks in animals and plants: a long
636 noncoding RNA perspective. Brief Funct Genomics. 2015 Mar;14(2):91–101.

637 34. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene
638 expression in primary immune cells identifies cell type-specific master regulators and roles of HLA
639 alleles. Nat Genet. 2012 Mar 25;44(5):502–10.

640 35. Sonawane AR, Platig J, Fagny M, Chen CY, Paulson JN, Lopes-Ramos CM, et al. Understanding
641 Tissue-Specific Gene Regulation. Cell Rep. 2017 Oct 24;21(4):1077–88.

642 36. Han L, Shen WJ, Bittner S, Kraemer FB, Azhar S. PPARs: regulators of metabolism and as
643 therapeutic targets in cardiovascular disease. Part II: PPAR-β/δ and PPAR-γ. Future Cardiol. 2017
644 May;13(3):279–96.

645 37. Kim J, Bagchi IC, Bagchi MK. Control of ovulation in mice by progesterone receptor-regulated gene
646 networks. Mol Hum Reprod. 2009 Dec;15(12):821–8.

647 38. Zhou J, Theesfeld CL, Yao K, Chen KM, Wong AK, Troyanskaya OG. Deep learning sequence-based
648 ab initio prediction of variant effects on expression and disease risk. Nat Genet. 2018
649 Aug;50(8):1171–9.

650 39. Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, et al. Effective gene
651 expression prediction from sequence by integrating long-range interactions. Nat Methods. 2021
652 Oct;18(10):1196–203.

653 40. Hudaiberdiev S, Taylor DL, Song W, Narisu N, Bhuiyan RM, Taylor HJ, et al. Modeling islet
654 enhancers using deep learning identifies candidate causal variants at loci associated with T2D and
655 glycemic traits. Proc Natl Acad Sci U S A. 2023 Aug 29;120(35):e2206612120.

656 41. Li S, Hannenhalli S, Ovcharenko I. De novo human brain enhancers created by single-nucleotide
657 mutations. Sci Adv. 2023 Feb 15;9(7):eadd2911.

658 42. Antczak M, Van Blerkom J. The vascular character of ovarian follicular granulosa cells: phenotypic
659 and functional evidence for an endothelial-like cell population. Hum Reprod Oxf Engl. 2000
660 Nov;15(11):2306–18.

661 43. Hartanti MD, Rosario R, Hummitzsch K, Bastian NA, Hatzirodos N, Bonner WM, et al. Could
662 perturbed fetal development of the ovary contribute to the development of polycystic ovary
663 syndrome in later life? PloS One. 2020;15(2):e0229351.

664 44. Joslin AC, Sobreira DR, Hansen GT, Sakabe NJ, Aneas I, Montefiori LE, et al. A functional genomics
665 pipeline identifies pleiotropy and cross-tissue effects within obesity-associated GWAS loci. Nat
666 Commun. 2021 Sep 6;12(1):5253.

667 45. Deng C, Whalen S, Steyert M, Ziffra R, Przytycki PF, Inoue F, et al. Massively parallel
668 characterization of regulatory elements in the developing human cortex. Science. 2024 May
669 24;384(6698):eadh0559.

670 46. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, et al. FoxA1 translates epigenetic
671 signatures into enhancer-driven lineage-specific transcription. Cell. 2008 Mar 21;132(6):958–70.

672 47. Mullen RD, Colvin SC, Hunter CS, Savage JJ, Walvoord EC, Bhangoo APS, et al. Roles of the LHX3
673 and LHX4 LIM-homeodomain factors in pituitary development. Mol Cell Endocrinol. 2007 Feb
674 1;265–266:190–5.

675 48. Topaloglu AK, Simsek E, Kocher MA, Mammadova J, Bober E, Kotan LD, et al. Inactivating NHLH2
676 variants cause idiopathic hypogonadotropic hypogonadism and obesity in humans. Hum Genet.
677 2022 Feb;141(2):295–304.

678 49. Cen C, Chen M, Zhou J, Zhang L, Duo S, Jiang L, et al. Inactivation of Wt1 causes pre-granulosa cell
679 to steroidogenic cell transformation and defect of ovary development†. Biol Reprod. 2020 Jun
680 23;103(1):60–9.

681 50. Li X, Wu X, Zhang H, Liu P, Xia L, Zhang N, et al. Analysis of single-cell RNA sequencing in human
682 oocytes with diminished ovarian reserve uncovers mitochondrial dysregulation and translation
683 deficiency. Reprod Biol Endocrinol. 2024 Nov 15;22(1):146.

684 51. Zhang Y, Federation AJ, Kim S, O'Keefe JP, Lun M, Xiang D, et al. Targeting nuclear receptor NR4A1-
685 dependent adipocyte progenitor quiescence promotes metabolic adaptation to obesity. J Clin
686 Invest. 2018 Nov 1;128(11):4898–911.

687 52. Nakanishi T, Tanaka R, Tonai S, Lee JY, Yamaoka M, Kawai T, et al. LH Induces De Novo Cholesterol
688 Biosynthesis via SREBP Activation in Granulosa Cells During Ovulation in Female Mice.
689 Endocrinology. 2021 Nov 1;162(11):bqab166.

690 53. Rajae TalbiTodd L StincicKaitlin FerrariChoi Ji HaeKarol WalecElizabeth MedveAchi GerutshangSilvia
691 LeónElizabeth A McCarthyOline K RønnekleivMartin J KellyVíctor M Navarro2024POMC neurons
692 control fertility through differential signaling of MC4R in Kisspeptin neurons. eLife13:RP100722.

693 54. Villa PA, Ruggiero-Ruff RE, Jamieson BB, Campbell RE, Coss D. Obesity Alters POMC and Kisspeptin
694 Neuron Cross Talk Leading to Reduced Luteinizing Hormone in Male Mice. J Neurosci Off J Soc
695 Neurosci. 2024 Jul 10;44(28):e0222242024.

696 55. Bohaczuk SC, Thackray VG, Shen J, Skowronska-Krawczyk D, Mellon PL. FSHB Transcription is
697 Regulated by a Novel 5' Distal Enhancer With a Fertility-Associated Single Nucleotide
698 Polymorphism. Endocrinology. 2021 Jan 1;162(1):bqaa181.

699 56. Liu XM, Yan MQ, Ji SY, Sha QQ, Huang T, Zhao H, et al. Loss of oocyte Rps26 in mice arrests oocyte
700 growth and causes premature ovarian failure. Cell Death Dis. 2018 Nov 19;9(12):1–15.

701 57. Svingen T, François M, Wilhelm D, Koopman P. Three-Dimensional Imaging of Prox1-EGFP
702 Transgenic Mouse Gonads Reveals Divergent Modes of Lymphangiogenesis in the Testis and Ovary.
703 PLOS ONE. 2012 Dec 20;7(12):e52620.

704 58. Karagiannopoulos A, Westholm E, Ofori JK, Cowan E, Esguerra JLS, Eliasson L. Glucocorticoid-
705 mediated induction of ZBTB16 affects insulin secretion in human islets and EndoC-βH1 β-cells.
706 iScience. 2023 May 19;26(5):106555.

707 59. Sun L, Ji S, Xie X, Si L, Liu S, Lin Y, et al. Deciphering the interaction between Twist1 and PPARγ
708 during adipocyte differentiation. Cell Death Dis. 2023 Nov 23;14(11):764.

709 60. Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-associated
710 variants within FTO form long-range functional connections with IRX3. Nature. 2014 Mar
711 20;507(7492):371–5.

712 61. Sobreira DR, Joslin AC, Zhang Q, Williamson I, Hansen GT, Farris KM, et al. Extensive pleiotropism
713 and allelic heterogeneity mediate metabolic effects of IRX3 and IRX5. Science. 2021 Jun
714 4;372(6546):1085–91.

715 62. Brynedal B, Choi J, Raj T, Bjornson R, Stranger BE, Neale BM, et al. Large-Scale trans-eQTLs Affect
716 Hundreds of Transcripts and Mediate Patterns of Transcriptional Co-regulation. Am J Hum Genet.
717 2017 Apr 6;100(4):581–91.

718 63. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO Obesity Variant
719 Circuitry and Adipocyte Browning in Humans. N Engl J Med. 2015 Sep 3;373(10):895–907.

720 64. Rotinen M, You S, Yang J, Coetzee SG, Reis-Sobreiro M, Huang WC, et al. ONECUT2 is a targetable
721 master regulator of lethal prostate cancer that suppresses the androgen axis. Nat Med. 2018
722 Dec;24(12):1887–98.

723 65. Zhao, ZH., Meng, TG., Chen, XY. et al. Spatiotemporal and single-cell atlases to dissect regional
724 specific cell types of the developing ovary. Commun Biol 8, 849 (2025).
725 https://doi.org/10.1038/s42003-025-08277-4

726 66. Ragvin A, Moro E, Fredman D, Navratilova P, Drivenes Ø, Engström PG, et al. Long-range gene
727 regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. Proc
728 Natl Acad Sci U S A. 2010 Jan 12;107(2):775–80.

729 67. Fu A, Koth ML, Brown RM, Shaw SA, Wang L, Krentz KJ, et al. IRX3 and IRX5 collaborate during
730 ovary development and follicle formation to establish responsive granulosa cells in the adult
731 mouse†. Biol Reprod. 2020 Aug 21;103(3):620–9.

732 68. Lamolet B, Pulichino AM, Lamonerie T, Gauthier Y, Brue T, Enjalbert A, et al. A pituitary cell-
733 restricted T box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins.
734 Cell. 2001 Mar 23;104(6):849–59.

735 69. Dai R, Sun Y. Altered GnRH neuron-glia networks close to interface of polycystic ovary syndrome:
736 Molecular mechanism and clinical perspectives. Life Sci. 2025 Jan 15;361:123318.

737 70. Corradin O, Saiakhova A, Akhtar-Zaidi B, Myeroff L, Willis J, Cowper-Sal lari R, et al. Combinatorial
738 effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to
739 confer susceptibility to common traits. Genome Res. 2014 Jan;24(1):1–13.

740 71. Kulakovskiy IV, Vorontsov IE, Yevshin IS, Sharipov RN, Fedorova AD, Rumynskiy EI, et al.
741 HOCOMOCO: towards a complete collection of transcription factor binding models for human and
742 mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res. 2018 Jan 4;46(D1):D252–9.

743 72. Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R, Castro-Mondragon JA, Ferenc K, Kumar V, et
744 al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding
745 profiles. Nucleic Acids Res. 2023 Nov 14;gkad1059.

746 73. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015 Jul
747 1;43(W1):W39–49.

748 74. Hudaiberdiev S, Taylor DL, Song W, Narisu N, Bhuiyan RM, Taylor HJ, et al. Modeling islet
749 enhancers using deep learning identifies candidate causal variants at loci associated with T2D and
750 glycemic traits. Proc Natl Acad Sci U S A. 2023 Aug 29;120(35):e2206612120.

751 75. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, et al. New developments on the
752 Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. 2020 Jan 8;48(D1):D882–
753 9.

754 76. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of
755 the Genome. Sci Rep. 2019 Jun 27;9(1):9354.

756 77. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics.
757 2011 Apr 1;27(7):1017–8.

758 78. Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R, Castro-Mondragon JA, Ferenc K, Kumar V, et
759 al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding
760 profiles. Nucleic Acids Res. 2024 Jan 5;52(D1):D174–82.

761 79. Vorontsov IE, Eliseeva IA, Zinkevich A, Nikonov M, Abramov S, Boytsov A, et al. HOCOMOCO in
762 2024: a rebuild of the curated collection of binding models for human and mouse transcription
763 factors. Nucleic Acids Res. 2024 Jan 5;52(D1):D154–63.

764 80. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for
765 Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007
766 Sep;81(3):559–75.

767 81. Hazell Pickering S, Abdelhalim M, Collas P, Briand N. Alternative isoform expression of key
768 thermogenic genes in human beige adipocytes. Front Endocrinol. 2024;15:1395750.

769 82. Weis-Banke SE, Lerdrup M, Kleine-Kohlbrecher D, Mohammad F, Sidoli S, Jensen ON, et al. Mutant
770 FOXL2C134W Hijacks SMAD4 and SMAD2/3 to Drive Adult Granulosa Cell Tumors. Cancer Res.
771 2020 Sep 1;80(17):3466–79.

772 83. Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of
773 large distributed datasets. Bioinformatics. 2010 Sep 1;26(17):2204–7.

774 84. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based Analysis of
775 ChIP-Seq (MACS). Genome Biol. 2008 Sep 17;9(9):R137.

776 85. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular
777 signatures database (MSigDB) 3.0. Bioinformatics. 2011 Jun 15;27(12):1739–40.

778 86. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in
779 2023: protein-protein association networks and functional enrichment analyses for any sequenced
780 genome of interest. Nucleic Acids Res. 2023 Jan 6;51(D1):D638–46.

781

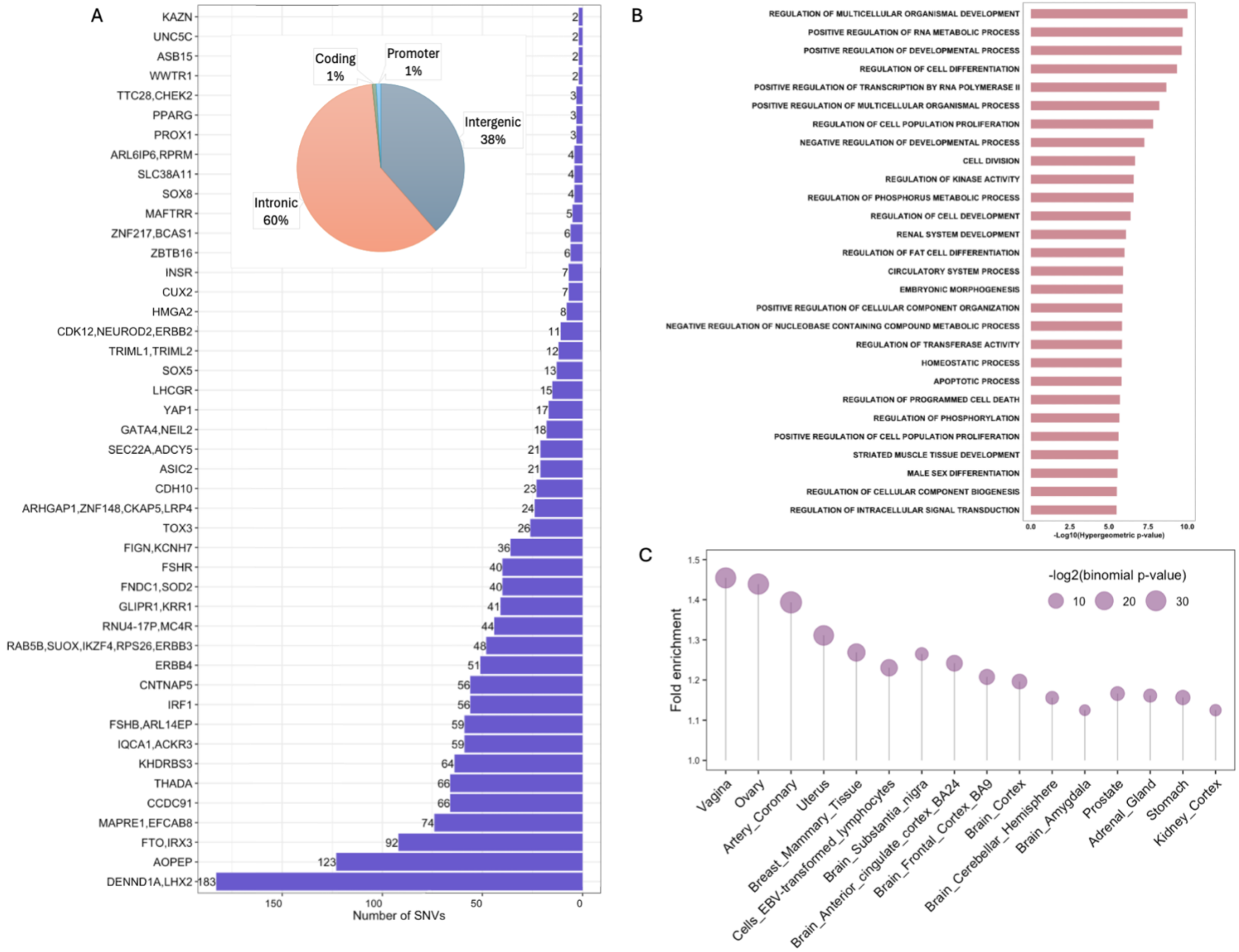

Figure 1: (A) PCOS susceptibility loci and their distribution in non-coding regions, (B) Gene Ontology annotations of target genes, (C) Fold enrichment of PCOS eVariants in GTEx cell-types (reported eVariants with enrichment binomial p-value < 0.01).

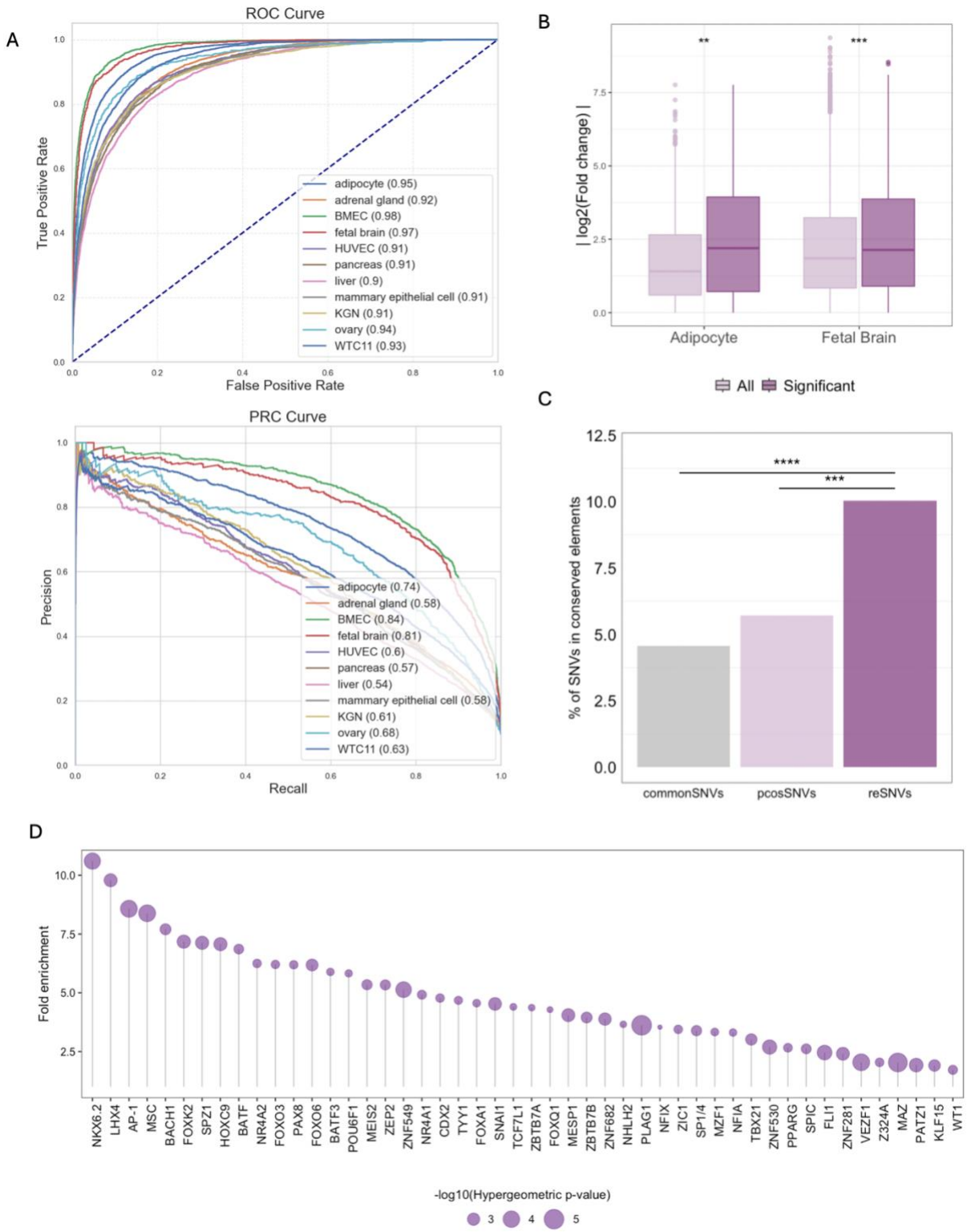

Figure 2: (A) ROC and PRC curves of eleven cell-type specific TREDNet models, (B) A comparison of fold change (alternate / reference allele) in TREDNet scores between all variants and those exhibiting significant change in enhancer activity in MPRA, using Wilcoxon test (C) Fraction of SNVs overlapping with phastCons elements conserved across 30 primates, (D) TFs enriched among reSNVs compared with control SNVs (hypergeometric p-value < 0.01). (ns: p > 0.05, *: p <= 0.05,**: p <= 0.01,***: p <= 0.001)

785

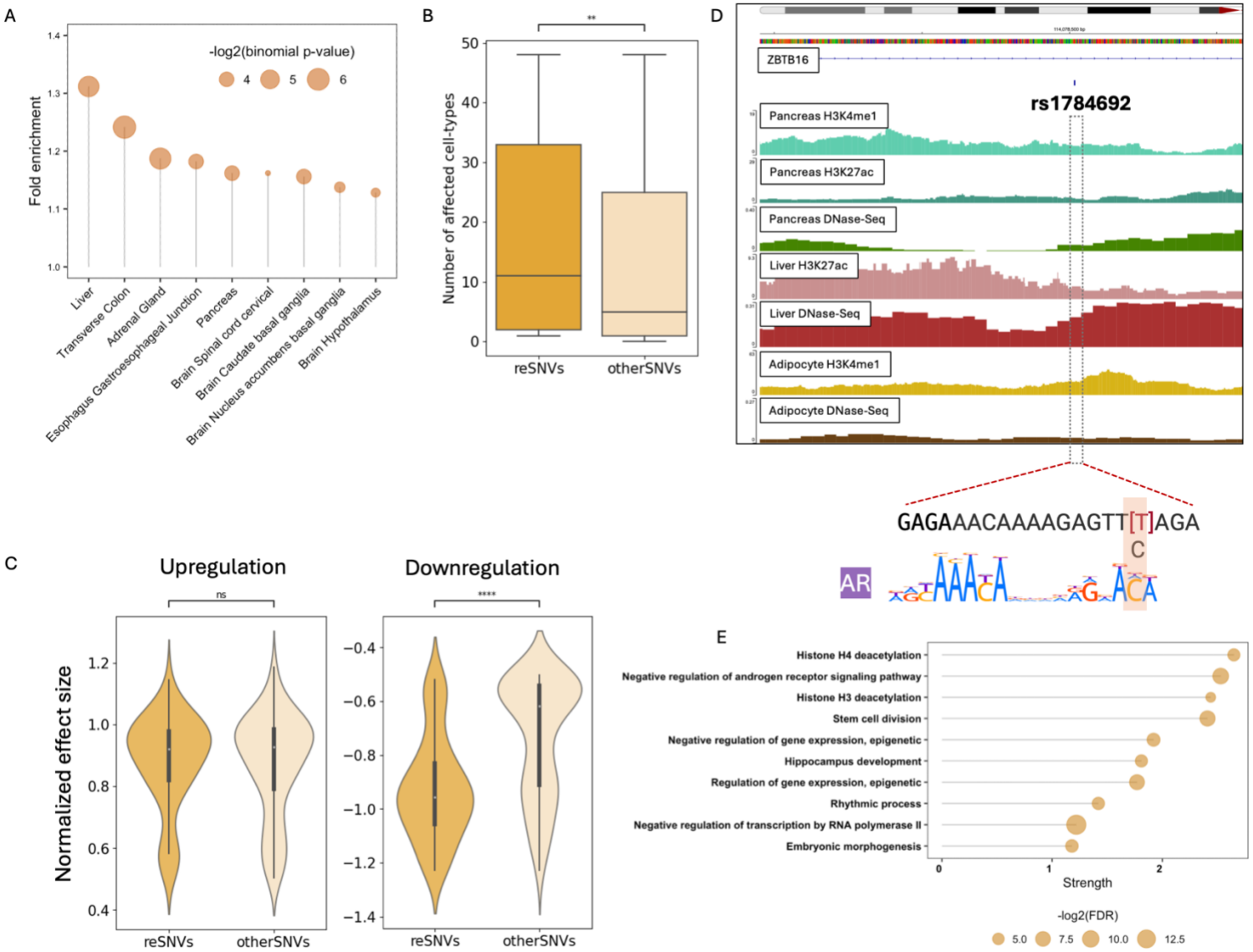

Figure 3: Regulatory impact of reSNVs prioritized by TREDNet. (A) Fold enrichment of reSNVs compared to pcosSNVs across cell-types (binomial $p < 0.01$). (B) Comparison of the number of GTEx cell-types impacted by reSNVs versus otherSNVs, (C) Normalized effect size of reSNVs versus otherSNVs. Left and right panels show differences for downregulating (NES ≤ –0.5) and upregulating (NES ≥ 0.5) variants, respectively. (D) Genomic overlap of an intronic reSNV (rs1784692) at the ZBTB16 locus with epigenomic features from cell types where it exhibits predicted allele-specific activity. The affected Androgen Receptor (AR) motif is shown below. (E) Functional enrichment of

biological processes in the ZBTB16 protein interaction network (STRING database). The plot shows the top 10 terms (FDR < 0.001), with enrichment strength calculated as $\log_{10}$(observed/expected). (ns: p > 0.05, *: p <= 0.05,**: p <= 0.01,***: p <= 0.001)

795 **Figure 4:** reSNVs in FTO locus exhibiting significant fold change in TREDNet predicted enhancer activity. (A) Overlap of reSNVs with active
796 regulatory regions of pathogenic cell-types (B) Intact Hi-C map of chromatin interactions from reSNVs in FTO locus in HUVEC
797 (doi:10.17989/ENCSR788FBI)