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) is the limited understanding of the molecular mechanisms driven by susceptibility loci, necessitating investigation into the regulatory pathways that contribute to the diverse phenotypic manifestations of PCOS. In this study, we integrated molecular and epigenomic annotations across proposed pathogenic cell types and employed a deep learning (DL) model to infer the cell-type-specific effects of risk variants. Our analysis revealed the role of these variants in brain and endocrine cell types affecting the binding sites of key transcription factors (TFs)—*FOXA1, FOXL1, WT1, SALL4,* and *CPEB1*—which regulate ovarian development, folliculogenesis, and steroid hormone signaling, contributing to disease-associated transcriptomic profiles. Our DL model, which is strongly correlated with MPRA data, identified enhancer-disrupting activity in 20% of the risk variants, particularly affecting TFs involved in androgen-mediated signaling, shedding light on the molecular consequences of hyperandrogenemia. Using the *IRX3-FTO* locus as a case study, we explored the potential cell-type-specific regulatory effects of risk variants in the fetal brain, pancreas, adipocytes, and an endothelial cell-line, which suggest that disruptions in IRX3 regulation (previously linked to obesity) may contribute to PCOS pathogenesis through diverse mechanisms, including neuronal development, metabolic regulation, and folliculogenesis. Our findings underscore the value of integrating DL models with epigenomic annotations to identify disease-relevant variants, explore the pleiotropic impact of disease risk loci, and gain novel insights into cross-cell-type regulatory interactions.

35 **Introduction**

36 Polycystic Ovary Syndrome (PCOS) is a multifactorial endocrine disorder characterized by
37 abnormal LH:FSH (Luteinizing hormone: Follicle Stimulating hormone) ratios and
38 elevated androgen levels, leading to anovulation, polycystic ovaries, and various
39 hyperandrogenism-related comorbidities [1–3]. The reproductive abnormalities in PCOS
40 stem from disruptions in the hypothalamic-pituitary-gonadal (HPG) axis, which also
41 contributes to other conditions like oligomenorrhoea, ovarian insufficiency, infertility,
42 hyper- and hypogonadism, and endometriosis [4]. This overlap in clinical features
43 complicates PCOS diagnosis, prompting the establishment of multiple diagnostic criteria
44 by the NIH, Rotterdam, and the Androgen Excess and PCOS Society. Based on a
45 consensus, diagnosis relies on clear indications of hyperandrogenism and ovulatory
46 dysfunction [3,5].
47
48 Decades of research on molecular mechanisms underlying the disease have identified
49 impaired steroidogenesis and folliculogenesis in theca and granulosa cells (GCs) as key
50 contributors to PCOS development [6]. These pathways are spatiotemporally regulated by
51 LH and FSH, secreted by the pituitary gland in response to hypothalamic Gonadotropin
52 Release Hormone (GnRH) based stimulation [7]. Moreover, PCOS often coincides with
53 hyperinsulinemia, though the molecular origins of the association between the two is
54 still being investigated. Hyperinsulinemia worsens hyperandrogenism by affecting
55 adrenal androgen production and reducing sex hormone-binding globulin (SHBG) levels
56 in the liver [8] and may also contribute to the metabolic co-morbidities such as obesity,
57 type-II diabetes and liver dysfunction [3,9].
58

59 The genetic basis of PCOS is thought to involve impaired regulation of the HPG axis [10].
60 Polymorphisms in genes coding for kisspeptin (Kiss1, an upstream regulator of GnRH),
61 GnRH receptor, Anti-Mullerian Hormone (AMH), LH, FSH, and their receptors have been
62 linked to impaired signaling in PCOS patients [11,12]. However, except for *AMHR*, no
63 functional studies directly connect these polymorphisms to the PCOS phenotype [13]. Twin
64 studies suggest an estimated 80% heritability [13], highlighting the need for detailed
65 investigations into the molecular mechanisms underlying PCOS etiology [14].
66
67 PCOS genome-wide association studies (GWAS) across diverse populations have
68 identified novel disease loci, including plausible candidates such as *FSHR, FSHB*, and
69 *LHCGR*, in addition to several others with no direct association with disease phenotypes
70 [10]. Variants in the loci *of THADA, DENND1A, IRF1, FTO* etc., have significant correlations
71 with the disease manifestation but have not been contextually studied. On the other
72 hand, polymorphisms in reproductive hormone receptors, including androgen (AR) and
73 estrogen receptors (*ESR1/2*), have been implicated in PCOS phenotypes [15,16]; however,
74 these associations have not been consistently replicated in GWAS studies. These
75 observations align with the omnigenic model of complex trait regulation, where core
76 genes directly influence the phenotype, while peripheral genes contribute through cell-
77 type-specific regulatory networks [17]. Variants exert their net impact through complex
78 genomic and epigenomic interactions, manifesting as GWAS association signals. This
79 highlights the need for further investigation into the regulatory networks governing
80 phenotypic complexity of PCOS. In this context, several isolated studies offer insights into
81 the involvement of GWAS-identified genes in PCOS pathophysiology [10]. For instance,
82 studies show that *ERBB4* and *GATA4* regulate folliculogenesis [18,19], *ZNF217* regulates
83 androgen production in theca cells [20], and *HMGA2* promotes granulosa cell proliferation
84 [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 the well-characterized regulatory locus of *IRX3*, we demonstrate how DL models combined with prior knowledge of key PCOS TFs can effectively prioritize causal variants.

Results

PCOS risk SNVs are enriched in neuroendocrine cell-types and influence gene expression across multiple cell-types

To conduct a comprehensive analysis of the regulatory features of PCOS risk loci, we identified 91 single nucleotide variants (SNVs) from twelve GWAS studies (Table S1). This set was expanded to 1,472 SNVs, referred herein as pcosSNVs, by including variants in linkage disequilibrium (LD) with GWAS-identified variants across all superpopulations (African, American, South and East Asian, European) with an $r^2 \geq 0.8$, obtained from SNiPA [23]. Adjacent variants were merged to define 50 genetic loci, named based on the nearest gene and/or previously associated genes in the literature (Figure 1A, Table S2). Most of

these variants are located within intronic regions, with the highest density observed in the *DENND1A* and *AOPEP* loci (Figure 1A). We then assigned target genes using a recent study that predicts enhancer-gene interactions across various cell types by integrating enhancer activity, 3D chromatin interactions, and DNase I hypersensitivity maps (Methods) [24]. This analysis yielded 97 target genes (Table S3), many of which were not previously linked to PCOS. These genes are significantly enriched in pathways related to cell development, differentiation, and apoptosis (hypergeometric p-value 10^{-6}, Figure 1B), highlighting their potential roles in the biological mechanisms underlying the five developmental stages of folliculogenesis in oocytes and granulosa cells [25].

We next investigated the functional impact of variants through their association with changes in gene expression characterized by the GTEx consortium [26]. Among 1,472 pcosSNVs, 832 overlapped with cis-eQTLs, termed eVariants (Table S4). Two-thirds of these pcosSNVs, including those in the RAB5B and IRF1 loci, are shared cell-type eQTLs, meaning they influence gene expression in multiple cell types. In contrast, the remaining variants, such as those in the *FSHR* and *ERBB4* loci, are cell-type specific (Figure S1a). The number of target genes scaled almost linearly with the number of affected cell types (Spearman correlation: 0.71, Figure S1b), suggesting that shared eQTLs may contribute to distinct cell-type specific regulatory networks by regulating different genes in different cell-types. Notably, eVariants in the locus of the GATA4 gene were linked to 39 genes across 49 cell types (Figure S1b), but all were located on the same chromosome and were not detected as trans-eQTLs by GTEx, preventing confirmation of their trans-regulatory role. However, 4% of eVariants targeted long non-coding RNAs (Table S4), suggesting their role as potential trans-eQTLs [27]. Meanwhile, the remaining 640 pcosSNVs that were not identified as GTEx eVariants belong to 18 susceptibility loci, including that of *CNTNAP5*,

135 *ASIC2*, and *CDH1*, (Table S5), likely due to low target gene expression or restricted
136 function in specific cell states in the dynamic transcription landscape that not captured
137 in bulk tissue analysis [26,28]. These pcosSNVs may play key spatiotemporal roles in
138 mediating GnRH response in the hypothalamus, pituitary regulation, and follicular phase
139 progression in PCOS [26].
140
141 We also examined the enrichment of PCOS eVariants across GTEx cell types. Compared
142 to a randomly selected set of 832 eQTLs (excluding PCOS eVariants), PCOS eVariants were
143 found to be enriched in brain cell types, reproductive hormone-producing tissues such as
144 the ovary and adrenal gland, as well as hormonally influenced tissues like the breast and
145 prostate (p-value < 0.001, Figure 1C). Since many PCOS eVariants are shared across cell-
146 types, they likely influence regulatory networks by mediating interactions between cell-
147 type-specific and ubiquitous transcription factors (TFs), thereby invoking cell-type-
148 specific regulatory pathways that contribute to distinct phenotypic outcomes in different
149 cellular contexts [29]. For example, *PPARG*, a susceptibility locus, plays a central role in
150 regulating lipid metabolism, adipocyte differentiation, gluconeogenesis, folliculogenesis,
151 and steroidogenesis through multiple TFs that are critical regulators of these biological
152 processes (www.kegg.jp/pathway/map=map03320) [30,31]. This suggests that causal
153 variants within this locus may contribute to distinct phenotypic outcomes through
154 pleiotropic effects across multiple cell types. Given these complexities, a deeper
155 understanding of specific TF functions within the PCOS regulatory landscape can help to
156 elucidate cell-type specific response in disease mechanisms.
157
158 **TFs acting as hormone response elements are enriched in pcosSNVs**

159 Next, we explored the prevalence and importance of TFs whose binding sites are
160 enriched for pcosSNVs. Specifically, we quantified the abundance of transcription factor
161 binding sites (TFBSs) affected by these variants and compared it to a control set of TFBSs
162 corresponding to 71,000 SNVs located within 100 kb of pcosSNVs (Methods). This
163 localized background enabled us to investigate the regulation of target genes within the
164 context of PCOS-specific biological processes, particularly for ubiquitously expressed
165 genes. Transcription factors (TFs) whose binding sites are enriched at pcosSNVs include
166 members of the forkhead family, notably *FOXA1*, which functions as a pioneer factor in
167 estrogen- and androgen-mediated responses [32], and *FOXL2*, a key ovarian transcription
168 factor associated with primary ovarian insufficiency [33] (p-value < 0.01, Figure 1D). Other
169 more abundant TFBSs include that of *EGR2* that regulates ovarian cell function and
170 survival [34,35], as well as *WT1* and *CPEB1* that regulate granulosa cell development and
171 proliferation [36,37]. Notably, pcosSNVs also exhibit a significant enrichment of *PROX1*
172 binding sites—an intriguing observation given that *PROX1* itself is a known susceptibility
173 locus for PCOS. This suggests a possible feedback regulatory mechanism involving *PROX1*
174 in PCOS pathophysiology. We also observed enrichment of TFs selectively expressed in
175 neuronal and ovarian cell types, such as *MNX1*, and *NKX6.1/2*, which, though not yet
176 linked to PCOS, are interesting candidates in neuroendocrine regulation.
177
178 In parallel, we examined the transcriptional basis of differentially expressed genes (DEGs)
179 between PCOS patients and healthy individuals in GCs—the reproductive hormone-
180 producing cell type that governs both folliculogenesis and steroidogenesis (Methods) [38].
181 Using the TF enrichment pipeline from the KnockTF database, which catalogs upstream
182 regulators of differentially expressed genes across various cell types (Methods) [39], we
183 identified 85 TFs whose knockdown or knockout results in differential expression of
184 target gene sets (Figure S2, Table S6). In addition to *WT1, FOXA1*, and *EGR2*, whose

185 binding sites are enriched in pcosSNVs, we also identified developmental regulators
186 *SALL4* and *NANOG*, which are essential for oocyte development [40], co-activators of
187 estrogen and androgen signaling, including *CCAR1/2* [41,42], and regulators of
188 steroidogenesis, such as *AHR, NR4A1, NR5A1 (SF1)* (p-value < 0.001, Figure S2). Using the
189 same background of 71k SNVs, we also assessed the enrichment of ChIP-Seq peaks within
190 pcosSNVs using data from extensively profiled cell lines HepG2, HEK293, WTC11, and
191 MCF-7 from ENCODE [24]. In HepG2, pcosSNVs were 40-fold more enriched for ChIP-Seq
192 peaks of *CCAR2* (p-value < 10^{-3}), Figure S3), which, along with *AGO2* and *RBFOX2*, was
193 also identified as an upstream regulator of DEGs in granulosa cells (Figure S2).
194 Interestingly, we also identified the enrichment of ChIP-Seq peaks of *NCOA2* in pcosSNVs,
195 another coactivator of *AR* and *ESR1*. We hypothesized that some of these TFs may be
196 involved in regulation through protein-protein interactions and not by direct binding at
197 consensus sites [43]. Overall, this analysis highlights key regulatory TFs involved in
198 folliculogenesis and endocrine signaling that may impact downstream pathways of
199 circulating reproductive hormone levels in PCOS.
200

201 **PcosSNVs of reproductive and metabolic subtype are enriched for distinct TFBSs and**
202 **phenotypic traits**

203 The phenotypic co-morbidities linked to PCOS, in addition to neuroendocrine
204 dysfunction, have led to the classification of two distinct subtypes—metabolic and
205 reproductive. The metabolic subtype is characterized by abnormal LH:FSH ratios, SHBG
206 levels, and/or elevated androgen levels, whereas the reproductive subtype is associated
207 with irregular BMI, glucose, and insulin levels. Based on the sub-phenotype associations
208 of GWAS susceptibility variants, we classified ten loci as metabolic and nineteen as
209 reproductive (Table S7). Of note, two separate variants within the *GATA4-NEIL2* locus

210 were linked to the reproductive and metabolic subtypes. We sought to investigate the
211 molecular mechanisms underlying these two phenotypes.
212
213 We first examined the differential enrichment of TFBSs in each subtype (Figure 2A,
214 Methods). The metabolic pcosSNVs showed significant enrichment for TFBSs of *TWIST1*
215 and *RXRB*, which regulate adipocyte differentiation [44,45], *PTF1A*, which is essential for
216 maintaining pancreatic cell identity [46], and steroid hormone nuclear receptors *RXRB*,
217 *RXRG* and *THRA* [47] (p-value < 0.01 vs reproductive pcosSNVs). Interestingly, binding sites
218 of TFs involved in the development and differentiation of GCs during folliculogenesis,
219 such as *SOX2*, *CPEB1*, and *FOXL2*, are enriched in the metabolic subtype. This suggests a
220 crucial role GCs, potentially through the regulation of reproductive hormone levels, in
221 driving metabolic abnormalities associated with PCOS. Conversely, the reproductive
222 subtype is enriched for TFBSs of *NR2F2*, *TBX2* and *NR1H4* (p-value < 0.01 vs metabolic
223 pcosSNVs), all of which have well-documented roles in steroidogenesis and
224 folliculogenesis [48,49].
225
226 The two subtypes are enriched for variants from distinct GWAS trait groups, correlating
227 with their phenotypic traits (Figure 2B, Methods). For example, reproductive subtype
228 variants exhibit more than 3-fold enrichment for endocrine traits like endometriosis and
229 uterine fibroids (p-value < 10^{-13} vs pcosSNVs), while metabolic subtype variants show
230 significant enrichment for obesity, BMI, and cholesterol (>5-fold, p-value < 10^{-30} vs
231 pcosSNVs). Furthermore, of the 249 metabolic and 462 reproductive variants, 203 (81%)
232 and 253 (54%) were identified as eQTLs in GTEx cell types. Compared to metabolic
233 eVariants, reproductive eVariants are enriched across a broader range of cell types,
234 whereas metabolic eVariants exhibit greater cell-type specificity, particularly in the
235 heart, adipose tissue, muscle, and esophagus (Figure 2C, Methods). The shared eQTLs

236 among reproductive eVariants suggest potential pleiotropic effects, while the cell-type-
237 specific eQTLs in the metabolic subtype likely exert targeted effects on their respective
238 cell types [50].
239
240 **Non-neuroendocrine pathogenic cell-types are proposed to manifest disease**
241 **phenotype in response to impaired hormone signaling**

242 The established role of the HPG axis in regulating circulating reproductive hormone levels
243 highlights the hypothalamus, pituitary, adrenal gland, and ovarian granulosa and theca
244 cells as key mediators of PCOS disease mechanisms. However, PCOS phenotypes extend
245 beyond these cell types, affecting the pancreas, adipocytes, liver, heart, and other
246 tissues, suggesting disruptions in downstream hormonal regulation within non-
247 neuroendocrine systems, in addition to potential pleiotropic effects of risk variants. We
248 thus categorized pathogenic cell-types into two groups: (a) primary pathogenic cell
249 types, which are responsible for synthesizing reproductive hormones, and (b) secondary
250 pathogenic cell types, which respond to these hormones through their respective
251 receptors [51].
252
253 Next, we analyzed the expression of receptors of reproductive hormones, namely,
254 *LHCGR, FSHR, AMHR, ESR1, ESR2* and *AR*, across different cell-types (Figure 3A) [3,16,52]. We
255 observed over 7-fold and 2-fold higher expression of *AR* in non-reproductive tissues such
256 as the liver and muscle, respectively, compared to its mean expression across all cell
257 types, possibly explaining the broader implications of hyperandrogenemia in PCOS
258 patients. Additionally, *AMHR2* is also expressed in the adrenal glands but its role in the
259 cellular context is not yet determined. Interestingly, *AR*-expressing cell types associate
260 with traits of the metabolic subtype (Figure 2B), suggesting that co-morbidities like

261 obesity, diabetes, and cardiovascular diseases [3] may arise from disrupted hormone
262 receptor networks. For instance, the role of androgens in regulating adipocyte
263 differentiation [53], and cholesterol biosynthesis in the liver [54], among other functions [55],
264 aligns with the expression of their receptor AR across non-neuroendocrine cell-types.
265 Additionally, the enrichment of ChIP-Seq peaks of their coactivators, such as *CCAR1/2*
266 and *NCOA*, among pcosSNVs in HepG2 (p-value < 0.001, Figure S3) further supports the
267 involvement of disrupted androgen signaling pathways in driving disease phenotypes in
268 non-neuroendocrine cell types. Together, these observations indicate that
269 understanding the regulatory networks of hormone receptors is essential to advance our
270 understanding of the molecular mechanisms underlying PCOS as disrupted circulating
271 hormone levels from primary pathogenic cell-types may influence other cell types
272 through altered hormone-responsive elements, further contributing to disease
273 pathology.
274
275 We used the set of enriched TFs in pcosSNVs and hormone receptors to obtain a protein
276 interaction network (Figure 3B). This network is enriched for pathways related to nuclear
277 receptors mediated hormone signaling, particularly by androgens, and adipogenesis, and
278 metabolism (Figure 3C), highlighting a complex regulatory framework within the HPG
279 axis governed by neuroendocrine hormones that may help in providing insights into the
280 co-morbidities associated with PCOS. Based on these observations, we propose a
281 signaling mechanism where dysregulated hormone biosynthesis in pathogenic cell types
282 disrupts hormone-responsive pathways in affected cell types (Figure 3D). Additionally,
283 pleiotropic effects of genetic variants in pathogenic cell types may influence diverse
284 cellular processes depending on the function of the target gene (Figure 3D), explaining
285 the multifaceted nature of PCOS.
286

287 We investigated plausible mechanisms that are disrupted in PCOS by analyzing the
288 epigenomic features of causal variants predicted at each susceptibility locus in both
289 pathogenic and affected cell-types using a deep learning algorithm, as described in the
290 following section.
291
292 **Prioritizing disease causal variants using a deep learning model**

293 The challenge of characterizing the cell-type-specific impact of thousands of
294 susceptibility variants in complex traits and diseases has led to the development of
295 computational approaches for inferring causal variants. DL models have been particularly
296 effective in predicting variant effects on gene regulation by integrating diverse cell-type-
297 specific epigenomic features [56,57]. We previously developed a convolutional neural
298 network based DL model, TREDNet, which can predict the effects of non-coding variants
299 on enhancer activity [58]. This two-phase DL model was demonstrated as a successful
300 approach in prioritizing causal variants of type 2 diabetes and autism [58,59]. Building on its
301 success, we applied TREDNet to investigate the regulatory mechanisms underlying PCOS.
302
303 We trained our model using cell-type-specific putative enhancers identified by co-
304 occurrence of the H3K27ac histone modification with DNase hypersensitive sites (DHS),
305 as proxies for active enhancers (Table S8, Methods). Aside from the pathogenic cell-
306 types, we included pancreas, liver, and adipocytes to investigate regulatory changes in
307 affected cell-types. Due to the lack of relevant epigenomic data from the pituitary and
308 hypothalamus, we included fetal brain instead, and brain microvascular endothelial cells
309 (BMEC), mammary epithelial cells and human umbilical vein endothelial cells (HUVEC) as
310 proxies for granulosa cells, given their similar H3K27ac profiles (Jaccard similarity index,
311 Table S9) and epithelial and endothelial-like properties [60]. Additionally, we incorporated

312 WTC11, a developmental cell line, to capture causal variants active during early
313 development, as fetal development has been implicated in PCOS onset later in life [61]. The
314 DL models demonstrated robust performance, achieving an area under the receiver
315 operating characteristic curve (auROC) ranging from 0.88 to 0.95 and an area under the
316 precision-recall curve (auPRC) ranging from 0.56 to 0.85 across the seven cell types
317 (Figure 4A).
318
319 To evaluate TREDNet's ability to identify causal variants, we examined the correlation
320 between TREDNet-predicted differences in allele-specific enhancer activity and those
321 determined through a massively parallel reporter assay (MPRA) in the developing human
322 brain and stem cell-derived adipocytes using our model trained on fetal brain and
323 adipocytes [62,63]. We compared allele-specific TREDNet scores across all assayed alleles
324 and those showing significant changes in reporter activity (Methods) and observed a
325 significantly higher fold change in TREDNet scores for the latter group (Mann-Whitney
326 test p-value = 0.001 for adipocytes and 10^{-5} for fetal brain, Figure 4B). These findings
327 highlight TREDNet's robustness in predicting causal variants across different cell types.
328
329 Next, we evaluated the impact of pcosSNVs within active regulatory regions across the
330 eleven selected cell-types, scoring both reference and risk alleles in all cell-type-specific
331 models (Methods). For pcosSNVs located in active regulatory regions marked by
332 H3K4me1, H3K27ac, or DNase/ATAC-Seq, we classified strengthening alleles as those
333 with scores below the threshold (determined at 10% FDR) for the reference allele and
334 above for the risk allele, while damaging alleles followed the opposite criterion. Applying
335 this approach, we identified 309 pcosSNVs with predicted allelic differences in activity,
336 termed reSNVs (Table S10). These reSNVs were significantly enriched in conserved
337 elements compared to both pcosSNVs and 13 million common SNVs from the 1000

338 Genomes catalog (binomial p-value = 0.0002 and 10^{-9}, respectively, Figure 4C).
339 Moreover, reSNVs exhibited greater enrichment in the brain, liver, adrenal gland, and
340 pancreas compared to pcosSNVs (Figure 4D), suggesting these as key cell-types impacted
341 by reSNVs.
342
343 To further assess the functional impact of reSNVs, we examined their influence on
344 transcription factor (TF) binding using allele-specific binding (ASB) data for ~400 TFs
345 across 557 cell types from the ADASTRA database (Methods) [64]. reSNVs showed a four-
346 fold enrichment in ASB SNVs compared to pcosSNVs (binomial p-value < 10^{-24}). Notably,
347 significantly affected TF binding sites included those for *FOXA1* and *ESR1* (Figure 4E),
348 suggesting the regulatory impact of reSNVs through hormone responsive elements.
349
350 Our deep learning-based approach identified 20% of the pcosSNVs as potential
351 regulatory variants, effectively narrowing down causal variants in loci such as *ERBB4*,
352 *LHCGR, MC4R*, etc (Figure S4). For example, among 51 variants in the *ERBB4* locus, we
353 identified rs79230362 as an enhancer-disrupting variant in HUVEC cells (Figure S5). This
354 variant is in LD with the GWAS SNP rs113168128 and is predicted to disrupt the binding
355 site of the *ELK1:SREBF2* motif complex (Figure S5). Given the established role of *SREBF2*
356 in steroidogenesis [65] and the highly cell-type-specific expression of ELK1 in granulosa cells
357 (Figure S5), this variant likely affects *ERBB4* expression, a key regulator of the oocyte
358 microenvironment during folliculogenesis [18]. Similarly, in the *MC4R* locus, we identified
359 rs17773430 as a causal enhancer-disrupting variant in WTC11 cells. *MC4R* is a critical
360 component of the melanocortin pathway and a well-established obesity susceptibility
361 gene that is also linked to PCOS. Notably, knockout studies of *MC4R* in mice result in both
362 obesity and infertility phenotypes, highlighting shared regulatory architectures
363 underlying these conditions [66]. rs17773430 is predicted to disrupt the binding site of

364 *TBX2/TBXT*, TFs responsible for the development of hypothalamus-pituitary axis [48]. Given
365 that reduced *MC4R* levels are associated with lower LH levels [67], this variant likely
366 contributes to PCOS etiology through its impact on HPG axis.
367
368 On the other hand, multiple reSNVs were identified in the locus of *DENND1A*, *FTO* and
369 *MAPRE1* (Figure S6). The significant overlap of reSNVs in *DENND1A* and *MAPRE1* locus
370 with active regulatory regions in the fetal brain and WTC11 suggests their potential role
371 in disease manifestation during early development. Notably, a reSNV in the *MAPRE1*
372 locus, rs187178, was validated as an enhancer-disrupting variant in the fetal brain and
373 functions as an eQTL for the neighboring gene *DNMT3B*, which regulates dynamic
374 methylation transitions during folliculogenesis [25]. In total, we identified 12 reSNVs
375 validated as enhancer disrupting variants in adipocytes and fetal brain through MPRA
376 studies (Table S11) [62,63].
377
378 Additionally, we gained insights into the potential role of certain loci whose involvement
379 in PCOS etiology is currently uncharacterized. For example, we identified a reSNV,
380 rs1784692, in the *ZBTB16* locus, which exhibited the highest predicted enhancer-
381 strengthening potential in the fetal brain, pancreas, adipocytes, WTC11, and liver (Figure
382 S7A). The T→C polymorphism enhances *AR* receptor binding (Figure S7B), suggesting a
383 possible association of this locus with cell-type-specific androgen response functions,
384 such as insulin secretion in the pancreas [68], and regulation of adipocyte differentiation
385 [45]. While *ZBTB16*'s role in PCOS remains unexplored, its protein interaction network is
386 enriched for androgen signaling (Figure S7C). These findings suggest that *ZBTB16* may
387 function as a susceptibility locus involved in androgen-mediated responses that are
388 disrupted in PCOS.
389

390 Of note, epigenomic data from fetal brain used by the DL model failed to capture the
391 regulatory impact of pathogenic variants in the *FSHB* locus, including rs10835638 and
392 rs11031006, which have been experimentally shown to reduce *FSHB* expression
393 restricted to the pituitary gland [69]. This underscores the necessity of incorporating
394 additional, relevant cell types for a more comprehensive study of the regulatory
395 landscape of PCOS, when experimental characterization of chromatin marks becomes
396 available for these cell types.
397

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

399 The regulatory locus within the intronic region of *FTO* is a well-known susceptibility locus
400 with significant implications in obesity and diabetes. Notably, it has been experimentally
401 validated to function as a distal enhancer of *IRX3*, a TF in PCOS-associated susceptibility
402 loci [70,71]. We hypothesized that this locus may have broader pleiotropic effects across
403 different cell types due to variations in the expression of *IRX3*, which may influence
404 multiple biological pathways [72]. Interestingly, the PCOS susceptibility variants localize in
405 the genomic region regulating *IRX3* (chr16:53731249–54975288) [73], suggesting that *IRX3*
406 is likely the target gene of the PCOS susceptibility locus as well (Figure S8).
407

408 We identified 12 reSNVs exhibiting significant fold changes across five cell types (Figure
409 S9). Among these, three variants— rs1421085, rs9940646 and rs9940128—have been
410 validated by MPRA studies to show allelic changes in enhancer activity in mouse
411 preadipocyte and neuronal cell lines [71], further supporting the predictive accuracy of
412 TREDNet in identifying causal variants. Interestingly, we predicted that rs1421085
413 additionally upregulates enhancer activity in BMEC by potentially modulating the binding

site of *ONECUT2* (Figure 5A), a suppressor of androgen receptor signaling which was recently identified as a marker of follicle growth [74,75].

Additionally, we identified another variant within the same locus, rs8050136, which is predicted as a causal variant in the pancreas and liver (Figure 5A, S8). This variant functions as an eQTL for *IRX3* in the pancreas, where IRX3 regulates the conversion of beta to epsilon cells, directly linking it to type 2 diabetes [76]. Notably, rs8050136 is also predicted to disrupt the binding site for *ONECUT1*, a transcription factor critical for pancreatic development (Figure 5). Together, these findings suggest that rs8050136 may serve as another causal variant for type 2 diabetes, possibly preferentially in PCOS patients.

To address the association of these variants with PCOS, we focused on a previous study that identified *IRX3* and the another gene in this susceptibility locus, *IRX5*, as key regulators of folliculogenesis in granulosa cells [77]. Using evidence from granulosa like cells, BMEC and HUVEC, we hypothesize that variants in this locus lead to impaired folliculogenesis, consequently disrupting androgen production in the pathogenic cell type—likely granulosa cells—through the dysregulated action of *IRX3/IRX5*. This disruption in androgen production may have pleiotropic effects on other cell types where these genes function within the androgen-responsive network, aligning with our proposed mechanism (Figure 3D). In this regard, rs9940128 emerges as a plausible causal variant as it forms chromatin contacts with promoters of *IRX3* and *IRX5* (Figure 5B) and is predicted to cause a significant fold change in enhancer activity in BMEC and HUVEC (Figure S9). Furthermore, the allelic effects of variants in this locus may also impact *IRX3/5*-mediated functions in hypothalamic neurons (Figure S8), as demonstrated in mice [71]. To explore this further, we analyzed the impact of these variants in fetal brain

440 and found that rs3751812 is located within binding sites of T-box family TFs (Figure 5).
441 Notably, members of the T-box family play a critical role in the commitment of
442 hypothalamus and pituitary lineages from neuronal precursors [48,78]. However, given the
443 short temporal window of expression of these TFs in neuronal development, inferring
444 causal mechanisms remains challenging. This highlights the necessity of using
445 epigenomic datasets across different developmental timepoints for a comprehensive
446 investigation.

447 The potential pleiotropic impact of disease-associated variants in non-pathogenic cell
448 types is often buffered by robust regulatory networks, preventing overt disease
449 manifestation. This suggests that assessing polygenic risk scores may be necessary to
450 fully understand their contribution to disease susceptibility. Given that variants in the
451 *FTO* locus have high minor allele frequencies (>0.4), which far exceed the prevalence of
452 PCOS, it is evident that the disease phenotypes emerge from the cumulative effects of
453 multiple dysregulated genes and pathways. Further investigations into polygenic
454 interactions and gene-environment influences will be essential to expand our
455 understanding of the complexity of PCOS.
456

457 **Discussion**

458 Our limited understanding of the regulatory landscape of PCOS stems from its complex
459 genetic architecture, which presents with heterogeneous phenotypes across different
460 cell types, individuals, and populations. This complexity has necessitated evolving
461 diagnostic criteria as our knowledge of the underlying pathophysiology expands. Several
462 key questions remain unresolved, including the genetic and molecular origins of
463 reproductive and metabolic dysfunction, the role of androgens and other hormones in
464 regulatory pathways, and the inheritance patterns affecting both males and females. To

465 date, GWAS have identified 50 genomic loci associated with PCOS across diverse
466 populations (Table S2). While the functional significance of genes such as *ERBB4*, *PPARG*,
467 and *IRX3* has been well established—leading to the use of their agonists as potential
468 treatments [18,31,77] —the precise molecular mechanisms remain elusive. Additionally,
469 advancements in whole-genome and exome sequencing continue to uncover novel loci,
470 further complicating our understanding of PCOS and highlighting the need for a deeper
471 exploration of the core regulatory mechanisms driving its pathophysiology.
472
473 Leveraging extensive genetic and epigenetic data, we sought to identify key mechanisms
474 linking PCOS susceptibility loci. We found that the susceptibility loci are enriched for
475 TFBSs associated with folliculogenesis, including *WT1*, *CPEB1*, and *FOXL1*. Subtype-
476 specific enrichment of distinct TFs provided deeper insights into the underlying biological
477 pathways. TFs such as *PTF1A*, *RXRB/G*, and *TWIST1*, which are involved in adipocyte and
478 pancreatic differentiation, were enriched in the metabolic subtype, while
479 steroidogenesis- and folliculogenesis-related TFs, including *NR2F2* and *NR1H4*, were
480 enriched in the reproductive subtype. By integrating these findings with an *in-silico*
481 causal variant assessment, we propose that neuroendocrine abnormalities in PCOS arise
482 from disruptions in hormone biosynthesis and/or their release from pathogenic cell-
483 types, manifesting as distinct phenotypes through regulatory interactions of their
484 receptors in affected cell-types. By prioritizing variants that disrupt PCOS associated TF
485 binding sites at susceptibility loci, we highlight the importance of TFs interacting with
486 hormone receptors—particularly androgens—as key modulators of PCOS-related
487 dysfunction. Our results highlight the need for further characterization of TFs such as the
488 *TBX* family, *MNX1*, etc., along with their interactions with hormonal receptors, to gain
489 deeper insights into PCOS regulatory biology.

490 Our prioritization of multiple variants at each susceptibility locus is based on the one-
491 gene multiple enhancers common assumption, according to which, the expression of a
492 target gene can be influenced by more than one variant [71,79]. For example, two distinct
493 variants in the *FSHB* locus—rs10835638 and rs11031006—alter *FSHB* expression,
494 ultimately contributing to infertility [69]. These variants may occur in different individuals,
495 leading to distinct, individual-specific phenotypes depending on the cell-type-specific
496 networks they modulate in a pleiotropic manner. However, a common theme among
497 reSNVs is their enrichment in reproductive hormone-responsive elements, such as
498 *FOXA1* and *ESR1* (Figure 4D), along with their interacting transcription factors, including
499 *ONECUT1/2*. This suggests that impaired downstream androgen signaling may be a key
500 mechanism contributing to PCOS pathophysiologys
501
502 The susceptibility loci of PCOS implicate genes such as *ZBTB16*, *AOPEP*, *THADA*, and
503 *CCDC91* (Figure 1A), which are ubiquitously expressed, raising the question of how
504 disease-specific variants selectively affect certain cell types. At the molecular level,
505 follicle progression involves signaling pathways like TGFβ, Hippo, Wnt, and mTOR, which
506 regulate fundamental processes such as cell proliferation, differentiation, and apoptosis
507 [7]. Why, then, do complex diseases manifest in only a subset of susceptible cell types? In
508 the case of *ZBTB16*, we predicted that rs1784692 strengthens enhancer activity by
509 increasing the binding affinity of *AR*, thereby implicating *ZBTB16* in downstream
510 pathways of androgen signaling. This suggests that perturbations in disease-relevant TF
511 interactions, specific to causal cell types, disrupt molecular networks in a way that
512 surpasses compensatory mechanisms in other cell types, thereby making certain cells
513 uniquely vulnerable. Consequently, transcription factors act as primary responders to
514 disease-associated alterations, preceding the genes they regulate, and may therefore
515 serve as more informative markers of disease susceptibility than the genes themselves.

516

517 Our analysis of the PCOS regulatory landscape reveals unifying molecular mechanisms
518 underlying disease phenotypes. However, a more comprehensive understanding of gene
519 regulatory networks requires integrating epigenomic datasets from key pathogenic cell
520 types—such as the pituitary gland, granulosa, and theca cells, and potentially, the
521 hypothalamus—across follicular phases to map the spatiotemporal regulation of genes
522 involved in steroidogenesis and folliculogenesis. Despite the hypothalamus's central role
523 in the HPG axis, regulatory networks mediated by GnRH signaling remain poorly
524 understood. Disruptions in this pathway may explain the involvement of risk genes such
525 as *CNTNAP5*, *ASIC2*, and *CUX2*, potentially linking PCOS to prevalent mental health
526 disorders [3]. Incorporating these datasets can enable the development of more inclusive
527 deep-learning models capable of predicting regulatory activity changes beyond enhancer
528 disruptions, offering deeper insights into PCOS pathophysiology.

529

530 Additionally, our PCOS subtype classification remains incomplete due to lack of data,
531 leaving some loci unassigned, which may exclude crucial transcription factors and
532 interactions essential for understanding regulatory networks. Lastly, our analysis of
533 causal variants was limited to those occurring within putative enhancers. However,
534 variants can impact gene regulation beyond enhancer activity. Variants located in
535 silencers or insulators may disrupt distal enhancer interactions, as observed with IRX3,
536 emphasizing the need for Hi-C data from pathogenic and affected cell types to resolve
537 target genes not identifiable through eQTL analysis. Lastly, a comprehensive approach
538 should also consider the trans-regulatory effects of risk variants—whether through TFs
539 encoded by susceptibility loci (*PROX1*, *SOX5/8*, *IRF1*) or non-coding RNAs that contribute
540 to epigenomic regulation of gene expression.

541

542 While preliminary, our results provide valuable insights into molecular mechanisms
543 underlying PCOS etiology. Future *in vitro* and *in vivo* characterization will be essential to
544 validate these predictions, potentially paving the way for novel, symptom-targeted
545 therapies for PCOS patients.

546 **Methods**

547 **PCOS susceptibility loci**

548 PCOS GWAS summary statistics were obtained from the NHGRI-GWAS catalog. Variants
549 in LD were expanded and clustered into 50 loci based on 100kb proximity. Loci subtypes
550 were assigned using GWAS-mapped traits and supporting literature (Table S7), then
551 translated to LD variants. The risk allele served as the alternate allele for GWAS variants,
552 while the minor allele was used for LD variants. All analyses were conducted using the
553 GRCh38 reference genome.

554

555 **Transcription factor binding sites**

556 Transcription factor binding site (TFBS) regions were defined by extending variant sites
557 by 30 bp on each side. TF binding profiles from HOCOMOCO [80] and JASPAR non
558 redundant collection [81] were analyzed using FIMO with default parameters [82]. Aside from
559 gain and loss of motifs, changes in motif scores were used to assess affinity differences
560 between reference and alternate alleles. A list of all the TFBSs gained, lost and modulated
561 for SNPs exhibiting significant fold change is provided in Table S10.

562 Differential enrichment of TFBSs between the metabolic and reproductive subtypes was
563 assessed using a binomial test, with TFBSs overlapping variants of one subtype analyzed
564 against the TFBSs of the other subtype as the background.

565

566 **Cell type specific DL models**

567 We used a two phase TREDNet model developed in our lab for cell-type specific enhancer
568 prediction [83]. The first phase of the model was pre-trained on 4560 genomic and
569 epigenomic profiles, which included DHS, ATAC-Seq, Histone ChIP-Seq and and TF ChIP-
570 Seq peaks from ENCODE v4 [84]. The second phase was fine-tuned to predict cell type
571 specific enhancers using training datasets described below. Chromosomes 8 and 9 were
572 held out for testing, chromosome 6 was used for validation and other autosomal
573 chromosomes were used to build the second phase model. The area under the ROC and
574 PRC curve for each of these models is listed in Table S9. The pre-trained phase-one model
575 has been deposited at https://doi.org/10.5281/zenodo.8161621.
576
577 Open chromatin (DHS or ATAC-Seq) and H3K27ac profiles for the causal cell-types were
578 downloaded from ENCODE [84] (Table S6). Positive datasets were defined as 2 kb regions
579 centered on overlaps between H3K27ac ChIP-seq peaks and chromatin accessibility peaks
580 of each cell type, excluding coding sequences, promoter proximal regions (<2kb from
581 TSS) and ENCODE blacklisted regions [85]. A 10-fold control dataset was generated for each
582 cell-type using randomly sampled 2kb fragments of the genome, excluding the positive
583 dataset of that cell type and blacklisted regions.
584
585 Each 2 kb fragment received an enhancer probability score. Active enhancers were
586 predicted at a 10% FPR with a 1:10 positive-to-control ratio. Variant effects were assessed
587 by scoring 2 kb regions centered on each variant for reference and alternate alleles. A
588 significant enhancer activity change was defined as an alternate/reference score ratio
589 >1.2 or <0.8.
590

591 **Enrichment analysis of TFBSs**

592 We used command line FIMO [86] to scan vertebrate TF motifs from JASPAR [87] and
593 HOCOMOCO [88] databases along the sequences, applying a p-value threshold of 10^{-5}.
594 Enrichment analysis used a background set of 1000 Genomes variants located within 100
595 kb of pcosSNVs but excluding those overlapping 30 bp flanks of pcosSNVs, exons and
596 promoter regions (UCSC-defined). Enrichment analysis was restricted to those TFBSs with
597 at least 10% occurrence relative to controls. Differential TFBS abundance in reproductive
598 vs. metabolic subtype variants was determined using the ratio of normalized TF binding
599 site counts between these enhancer categories.

600

601 **Data and tools**

602 The H3K27ac peaks for KGN cells and adipocytes were sourced from literature [89,90]. The
603 KGN wig file was converted to NarrowPeak format using UCSC BigWig tools [91] and MACS
604 peak calling software [92].
605 Motif logos were retrieved from HOCOMOCO database [88]. Ontology enrichment of
606 pcosSNVs was performed using the Molecular Signatures Database [93]. Protein interaction
607 networks and enriched pathways (Figure 3C) were obtained from STRING database [94].
608 Evolutionary conservation of genomic regions was measured by their extent of overlap
609 with phastCons elements conserved across 30 primates
610 (https://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/
611 phastConsElements30way.txt.gz).

612

613 **Declarations**

614

615 **Acknowledgements**

This research was supported by the Division of Intramural Research of the National Library of Medicine (NLM), National Institutes of Health. This work utilized the computational resources of the NIH HPC Biowulf cluster.

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.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Please see the section "Data and tools" and supplementary tables.

Ethics approval and consent to participate.

Not applicable.

Consent for publication

Not applicable.

1. References

1. Joham, A. E. *et al.* Polycystic ovary syndrome. *Lancet Diabetes Endocrinol.* **10**, 668–680 (2022).
2. Palomba, S., Piltonen, T. T. & Giudice, L. C. Endometrial function in women with polycystic ovary syndrome: a comprehensive review. *Hum. Reprod. Update* **27**, 584–618 (2021).
3. Stener-Victorin, E. *et al.* Polycystic ovary syndrome. *Nat. Rev. Dis. Primer* **10**, 1–23 (2024).
4. Zaidi, M., Yuen, T. & Kim, S.-M. Pituitary crosstalk with bone, adipose tissue and brain. *Nat. Rev. Endocrinol.* **19**, 708–721 (2023).
5. McCartney, C. R. & Marshall, J. C. Polycystic Ovary Syndrome. *N. Engl. J. Med.* **375**, 54–64 (2016).
6. Abedel-Majed, M. A., Romereim, S. M., Davis, J. S. & Cupp, A. S. Perturbations in Lineage Specification of Granulosa and Theca Cells May Alter Corpus Luteum Formation and Function. *Front. Endocrinol.* **10**, (2019).
7. Wang, K. & Li, Y. Signaling pathways and targeted therapeutic strategies for polycystic ovary syndrome. *Front. Endocrinol.* **14**, 1191759 (2023).
8. Liao, B., Qiao, J. & Pang, Y. Central Regulation of PCOS: Abnormal Neuronal-Reproductive-Metabolic Circuits in PCOS Pathophysiology. *Front. Endocrinol.* **12**, 667422 (2021).
9. Zhang, C.-H., Liu, X.-Y. & Wang, J. Essential Role of Granulosa Cell Glucose and Lipid Metabolism on Oocytes and the Potential Metabolic Imbalance in Polycystic Ovary Syndrome. *Int. J. Mol. Sci.* **24**, 16247 (2023).
10. Dapas, M. & Dunaif, A. Deconstructing a Syndrome: Genomic Insights Into PCOS Causal Mechanisms and Classification. *Endocr. Rev.* **43**, 927–965 (2022).

658 11. Gorsic, L. K., Dapas, M., Legro, R. S., Hayes, M. G. & Urbanek, M. Functional Genetic Variation in
659 the Anti-Müllerian Hormone Pathway in Women With Polycystic Ovary Syndrome. *J. Clin.*
660 *Endocrinol. Metab.* **104**, 2855–2874 (2019).
661 12. Caburet, S. *et al.* A homozygous mutation of GNRHR in a familial case diagnosed with polycystic
662 ovary syndrome. *Eur. J. Endocrinol.* **176**, K9–K14 (2017).
663 13. Chen, Y. *et al.* Genetic and Epigenetic Landscape for Drug Development in Polycystic Ovary
664 Syndrome. *Endocr. Rev.* **45**, 437–459 (2024).
665 14. Vink, J. M., Sadrzadeh, S., Lambalk, C. B. & Boomsma, D. I. Heritability of polycystic ovary
666 syndrome in a Dutch twin-family study. *J. Clin. Endocrinol. Metab.* **91**, 2100–2104 (2006).
667 15. Wang, F. *et al.* Alternative splicing of the androgen receptor in polycystic ovary syndrome. *Proc.*
668 *Natl. Acad. Sci. U. S. A.* **112**, 4743–4748 (2015).
669 16. Muccee, F. *et al.* Exploring the association of ESR1 and ESR2 gene SNPs with polycystic ovary
670 syndrome in human females: a comprehensive association study. *J. Ovarian Res.* **17**, 27 (2024).
671 17. Mathieson, I. The omnigenic model and polygenic prediction of complex traits. *Am. J. Hum. Genet.*
672 **108**, 1558–1563 (2021).
673 18. Veikkolainen, V. *et al.* Erbb4 regulates the oocyte microenvironment during folliculogenesis. *Hum.*
674 *Mol. Genet.* **29**, 2813–2830 (2020).
675 19. LaVoie, H. A. The GATA-keepers of ovarian development and folliculogenesis. *Biol. Reprod.* **91**, 38
676 (2014).
677 20. Waterbury, J. S. *et al.* The PCOS GWAS Candidate Gene ZNF217 Influences Theca Cell Expression of
678 DENND1A.V2, CYP17A1, and Androgen Production. *J. Endocr. Soc.* **6**, bvac078 (2022).
679 21. Li, M. *et al.* The HMGA2-IMP2 Pathway Promotes Granulosa Cell Proliferation in Polycystic Ovary
680 Syndrome. *J. Clin. Endocrinol. Metab.* **104**, 1049–1059 (2019).

681 22. Xi, Y. *et al.* HMGA2 promotes adipogenesis by activating C/EBPβ-mediated expression of PPARγ.
682 *Biochem. Biophys. Res. Commun.* **472**, 617–623 (2016).
683 23. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmüller, G. SNiPA: an interactive, genetic
684 variant-centered annotation browser. *Bioinforma. Oxf. Engl.* **31**, 1334–1336 (2015).
685 24. Gschwind, A. R. *et al.* An encyclopedia of enhancer-gene regulatory interactions in the human
686 genome. *bioRxiv* 2023.11.09.563812 (2023) doi:10.1101/2023.11.09.563812.
687 25. Zhang, Y. *et al.* Transcriptome Landscape of Human Folliculogenesis Reveals Oocyte and Granulosa
688 Cell Interactions. *Mol. Cell* **72**, 1021-1034.e4 (2018).
689 26. GTEx Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–
690 213 (2017).
691 27. Bai, Y., Dai, X., Harrison, A. P. & Chen, M. RNA regulatory networks in animals and plants: a long
692 noncoding RNA perspective. *Brief. Funct. Genomics* **14**, 91–101 (2015).
693 28. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific
694 master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).
695 29. Sonawane, A. R. *et al.* Understanding Tissue-Specific Gene Regulation. *Cell Rep.* **21**, 1077–1088
696 (2017).
697 30. Han, L., Shen, W.-J., Bittner, S., Kraemer, F. B. & Azhar, S. PPARs: regulators of metabolism and as
698 therapeutic targets in cardiovascular disease. Part II: PPAR-β/δ and PPAR-γ. *Future Cardiol.* **13**,
699 279–296 (2017).
700 31. Kim, J., Bagchi, I. C. & Bagchi, M. K. Control of ovulation in mice by progesterone receptor-
701 regulated gene networks. *Mol. Hum. Reprod.* **15**, 821–828 (2009).
702 32. Lupien, M. *et al.* FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific
703 transcription. *Cell* **132**, 958–970 (2008).

704 33. Georges, A. *et al.* FOXL2: a central transcription factor of the ovary. *J. Mol. Endocrinol.* **52**, R17-33
705 (2014).
706 34. Lee, S. L. *et al.* Luteinizing hormone deficiency and female infertility in mice lacking the
707 transcription factor NGFI-A (Egr-1). *Science* **273**, 1219–1221 (1996).
708 35. Jin, H. *et al.* EGR2 is a gonadotropin-induced survival factor that controls the expression of IER3 in
709 ovarian granulosa cells. *Biochem. Biophys. Res. Commun.* **482**, 877–882 (2017).
710 36. Cen, C. *et al.* Inactivation of Wt1 causes pre-granulosa cell to steroidogenic cell transformation and
711 defect of ovary development†. *Biol. Reprod.* **103**, 60–69 (2020).
712 37. Xuan, F. *et al.* CPEB1 induces autophagy and promotes apoptosis in ovarian granulosa cells of
713 polycystic ovary syndrome. *Mol. Reprod. Dev.* **91**, e23741 (2024).
714 38. Li, J. *et al.* Molecular Features of Polycystic Ovary Syndrome Revealed by Transcriptome Analysis of
715 Oocytes and Cumulus Cells. *Front. Cell Dev. Biol.* **9**, 735684 (2021).
716 39. Feng, C. *et al.* KnockTF 2.0: a comprehensive gene expression profile database with
717 knockdown/knockout of transcription (co-)factors in multiple species. *Nucleic Acids Res.* **52**, D183–
718 D193 (2024).
719 40. Overland, M. R. *et al.* Development of the human ovary: Fetal through pubertal ovarian
720 morphology, folliculogenesis and expression of cellular differentiation markers. *Differ. Res. Biol.*
721 *Divers.* **129**, 37–59 (2023).
722 41. Trauernicht, A. M., Kim, S. J., Kim, N. H. & Boyer, T. G. Modulation of estrogen receptor alpha
723 protein level and survival function by DBC-1. *Mol. Endocrinol. Baltim. Md* **21**, 1526–1536 (2007).
724 42. Fu, J. *et al.* Deleted in breast cancer 1, a novel androgen receptor (AR) coactivator that promotes
725 AR DNA-binding activity. *J. Biol. Chem.* **284**, 6832–6840 (2009).

726 43. Jindal, G. A. & Farley, E. K. Enhancer grammar in development, evolution, and disease:
727 dependencies and interplay. *Dev. Cell* **56**, 575–587 (2021).
728 44. Siersbæk, R. *et al.* Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte
729 Differentiation. *Mol. Cell* **66**, 420-435.e5 (2017).
730 45. Sun, L. *et al.* Deciphering the interaction between Twist1 and PPARγ during adipocyte
731 differentiation. *Cell Death Dis.* **14**, 764 (2023).
732 46. Weedon, M. N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic
733 agenesis. *Nat. Genet.* **46**, 61–64 (2014).
734 47. Evans, R. M. & Mangelsdorf, D. J. Nuclear Receptors, RXR, and the Big Bang. *Cell* **157**, 255–266
735 (2014).
736 48. Pontecorvi, M., Goding, C. R., Richardson, W. D. & Kessaris, N. Expression of Tbx2 and Tbx3 in the
737 developing hypothalamic-pituitary axis. *Gene Expr. Patterns GEP* **8**, 411–417 (2008).
738 49. Candelaria, N. R. & Richards, J. S. Targeted deletion of NR2F2 and VCAM1 in theca cells impacts
739 ovarian follicular development: insights into polycystic ovary syndrome?†. *Biol. Reprod.* **110**, 782–
740 797 (2024).
741 50. Gamazon, E. R. *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex
742 disease- and trait-associated variation. *Nat. Genet.* **50**, 956–967 (2018).
743 51. Concannon, P., Rich, S. S. & Nepom, G. T. Genetics of type 1A diabetes. *N. Engl. J. Med.* **360**, 1646–
744 1654 (2009).
745 52. Wu, P.-F. *et al.* Polycystic ovary syndrome is causally associated with estrogen receptor-positive
746 instead of estrogen receptor-negative breast cancer: a Mendelian randomization study. *Am. J.
747 Obstet. Gynecol.* **223**, 583–585 (2020).

748 53. Chazenbalk, G. *et al.* Androgens inhibit adipogenesis during human adipose stem cell commitment
749 to preadipocyte formation. *Steroids* **78**, 920–926 (2013).
750 54. Krycer, J. R. & Brown, A. J. Cross-talk between the androgen receptor and the liver X receptor:
751 implications for cholesterol homeostasis. *J. Biol. Chem.* **286**, 20637–20647 (2011).
752 55. Davey, R. A. & Grossmann, M. Androgen Receptor Structure, Function and Biology: From Bench to
753 Bedside. *Clin. Biochem. Rev.* **37**, 3–15 (2016).
754 56. Zhou, J. *et al.* Deep learning sequence-based ab initio prediction of variant effects on expression
755 and disease risk. *Nat. Genet.* **50**, 1171–1179 (2018).
756 57. Avsec, Ž. *et al.* Effective gene expression prediction from sequence by integrating long-range
757 interactions. *Nat. Methods* **18**, 1196–1203 (2021).
758 58. Hudaiberdiev, S. *et al.* Modeling islet enhancers using deep learning identifies candidate causal
759 variants at loci associated with T2D and glycemic traits. *Proc. Natl. Acad. Sci. U. S. A.* **120**,
760 e2206612120 (2023).
761 59. Li, S., Hannenhalli, S. & Ovcharenko, I. De novo human brain enhancers created by single-
762 nucleotide mutations. *Sci. Adv.* **9**, eadd2911 (2023).
763 60. Antczak, M. & Van Blerkom, J. The vascular character of ovarian follicular granulosa cells:
764 phenotypic and functional evidence for an endothelial-like cell population. *Hum. Reprod. Oxf. Engl.*
765 **15**, 2306–2318 (2000).
766 61. Hartanti, M. D. *et al.* Could perturbed fetal development of the ovary contribute to the
767 development of polycystic ovary syndrome in later life? *PloS One* **15**, e0229351 (2020).
768 62. Joslin, A. C. *et al.* A functional genomics pipeline identifies pleiotropy and cross-tissue effects
769 within obesity-associated GWAS loci. *Nat. Commun.* **12**, 5253 (2021).

770 63. Deng, C. *et al.* Massively parallel characterization of regulatory elements in the developing human
771 cortex. *Science* **384**, eadh0559 (2024).
772 64. Abramov, S. *et al.* Landscape of allele-specific transcription factor binding in the human genome.
773 *Nat. Commun.* **12**, 2751 (2021).
774 65. Nakanishi, T. *et al.* LH Induces De Novo Cholesterol Biosynthesis via SREBP Activation in Granulosa
775 Cells During Ovulation in Female Mice. *Endocrinology* **162**, bqab166 (2021).
776 66. Talbi, R. *et al.* POMC neurons control fertility through differential signaling of MC4R in Kisspeptin
777 neurons. *BioRxiv Prepr. Serv. Biol.* 2024.02.18.580873 (2024) doi:10.1101/2024.02.18.580873.
778 67. Villa, P. A., Ruggiero-Ruff, R. E., Jamieson, B. B., Campbell, R. E. & Coss, D. Obesity Alters POMC and
779 Kisspeptin Neuron Cross Talk Leading to Reduced Luteinizing Hormone in Male Mice. *J. Neurosci.*
780 *Off. J. Soc. Neurosci.* **44**, e0222242024 (2024).
781 68. Karagiannopoulos, A. *et al.* Glucocorticoid-mediated induction of ZBTB16 affects insulin secretion
782 in human islets and EndoC-βH1 β-cells. *iScience* **26**, 106555 (2023).
783 69. Bohaczuk, S. C., Thackray, V. G., Shen, J., Skowronska-Krawczyk, D. & Mellon, P. L. FSHB
784 Transcription is Regulated by a Novel 5' Distal Enhancer With a Fertility-Associated Single
785 Nucleotide Polymorphism. *Endocrinology* **162**, bqaa181 (2021).
786 70. Smemo, S. *et al.* Obesity-associated variants within FTO form long-range functional connections
787 with IRX3. *Nature* **507**, 371–375 (2014).
788 71. Sobreira, D. R. *et al.* Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of
789 IRX3 and IRX5. *Science* **372**, 1085–1091 (2021).
790 72. Brynedal, B. *et al.* Large-Scale trans-eQTLs Affect Hundreds of Transcripts and Mediate Patterns of
791 Transcriptional Co-regulation. *Am. J. Hum. Genet.* **100**, 581–591 (2017).

73. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N. Engl. J. Med.* **373**, 895–907 (2015).

74. Rotinen, M. *et al.* ONECUT2 is a targetable master regulator of lethal prostate cancer that suppresses the androgen axis. *Nat. Med.* **24**, 1887–1898 (2018).

75. Zhao, Z.-H., Meng, T.-G., Gao, F., Schatten, H. & Sun, Q.-Y. Spatiotemporal and single-cell atlases to dissect cell lineage differentiation and regional specific cell types in mouse ovary morphogenesis. 2023.07.21.549985 Preprint at https://doi.org/10.1101/2023.07.21.549985 (2023).

76. Ragvin, A. *et al.* Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 775–780 (2010).

77. Fu, A. *et al.* IRX3 and IRX5 collaborate during ovary development and follicle formation to establish responsive granulosa cells in the adult mouse†. *Biol. Reprod.* **103**, 620–629 (2020).

78. Lamolet, B. *et al.* A pituitary cell-restricted T box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. *Cell* **104**, 849–859 (2001).

79. Corradin, O. *et al.* Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* **24**, 1–13 (2014).

80. Kulakovskiy, I. V. *et al.* HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).

81. Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* gkad1059 (2023) doi:10.1093/nar/gkad1059.

82. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).

815 83. Hudaiberdiev, S. *et al.* Modeling islet enhancers using deep learning identifies candidate causal
816 variants at loci associated with T2D and glycemic traits. *Proc. Natl. Acad. Sci. U. S. A.* **120**,
817 e2206612120 (2023).
818 84. Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data portal.
819 *Nucleic Acids Res.* **48**, D882–D889 (2020).
820 85. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic
821 Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).
822 86. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif.
823 *Bioinformatics* **27**, 1017–1018 (2011).
824 87. Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription
825 factor binding profiles. *Nucleic Acids Res.* **52**, D174–D182 (2024).
826 88. Vorontsov, I. E. *et al.* HOCOMOCO in 2024: a rebuild of the curated collection of binding models for
827 human and mouse transcription factors. *Nucleic Acids Res.* **52**, D154–D163 (2024).
828 89. Hazell Pickering, S., Abdelhalim, M., Collas, P. & Briand, N. Alternative isoform expression of key
829 thermogenic genes in human beige adipocytes. *Front. Endocrinol.* **15**, 1395750 (2024).
830 90. Weis-Banke, S. E. *et al.* Mutant FOXL2C134W Hijacks SMAD4 and SMAD2/3 to Drive Adult
831 Granulosa Cell Tumors. *Cancer Res.* **80**, 3466–3479 (2020).
832 91. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling
833 browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
834 92. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
835 93. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740
836 (2011).

837 94. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and
838 functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**,
839 D638–D646 (2023).
840

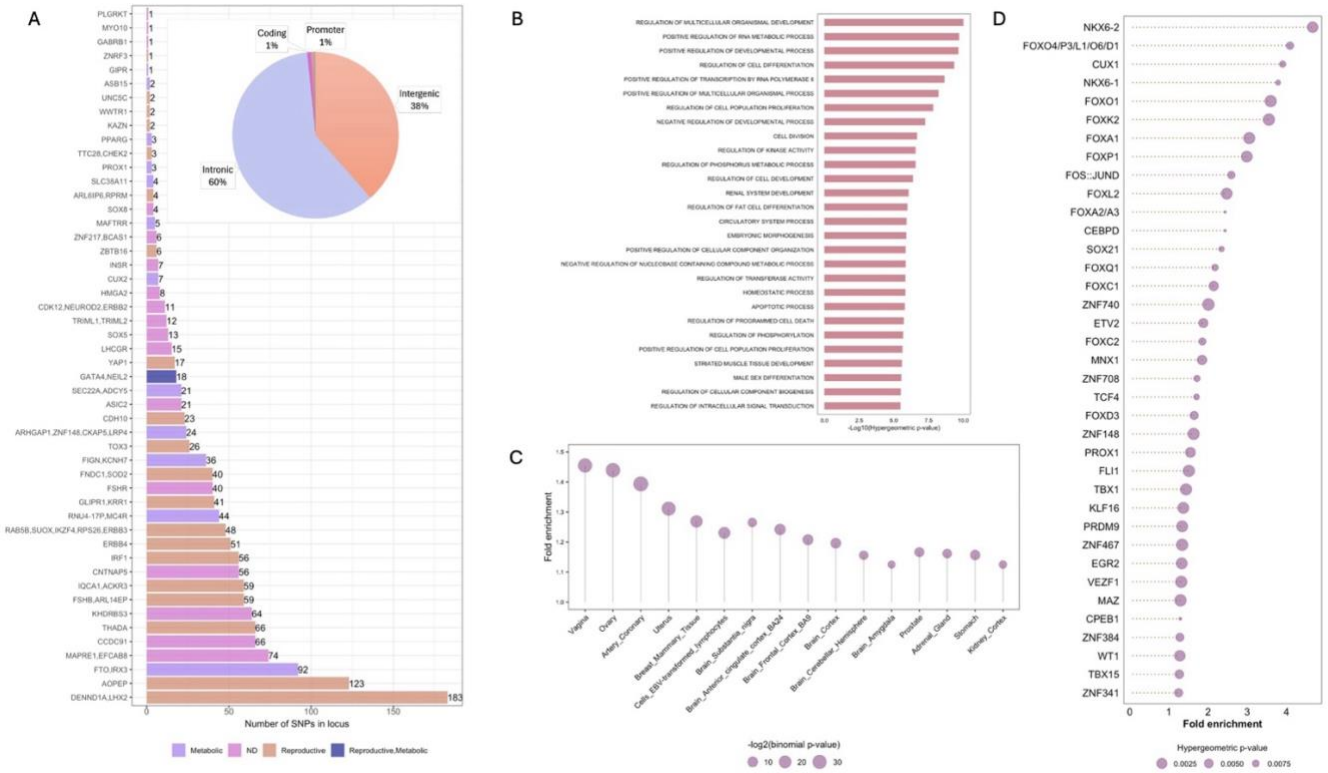

Figure 1: (A) PCOS susceptibility loci and their distribution in non-coding regions, (B) Gene Ontology annotations of target genes, (C) Fold enrichment of PCOS eVariants in GTEx cell-types (p-value < 0.01), (D) Fold enrichment of TFBSs in pcosSNVs compared with other variants in the susceptibility loci.

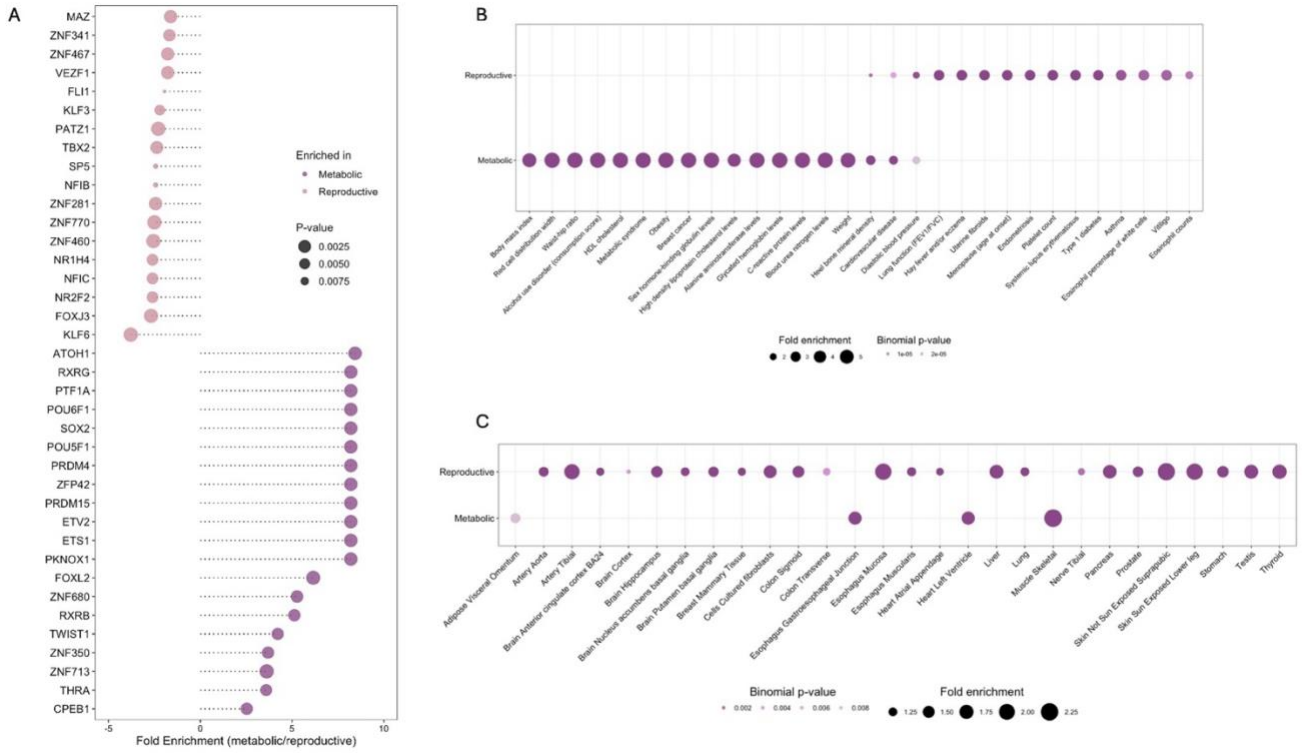

Figure 2: (A) Variation in TFBS enrichment between metabolic and reproductive subtypes, (B) GWAS traits associated with each subtype, (C) GTEx eQTL enrichment across the two subtypes. All reported terms meet the significance threshold of p-value < 0.01.

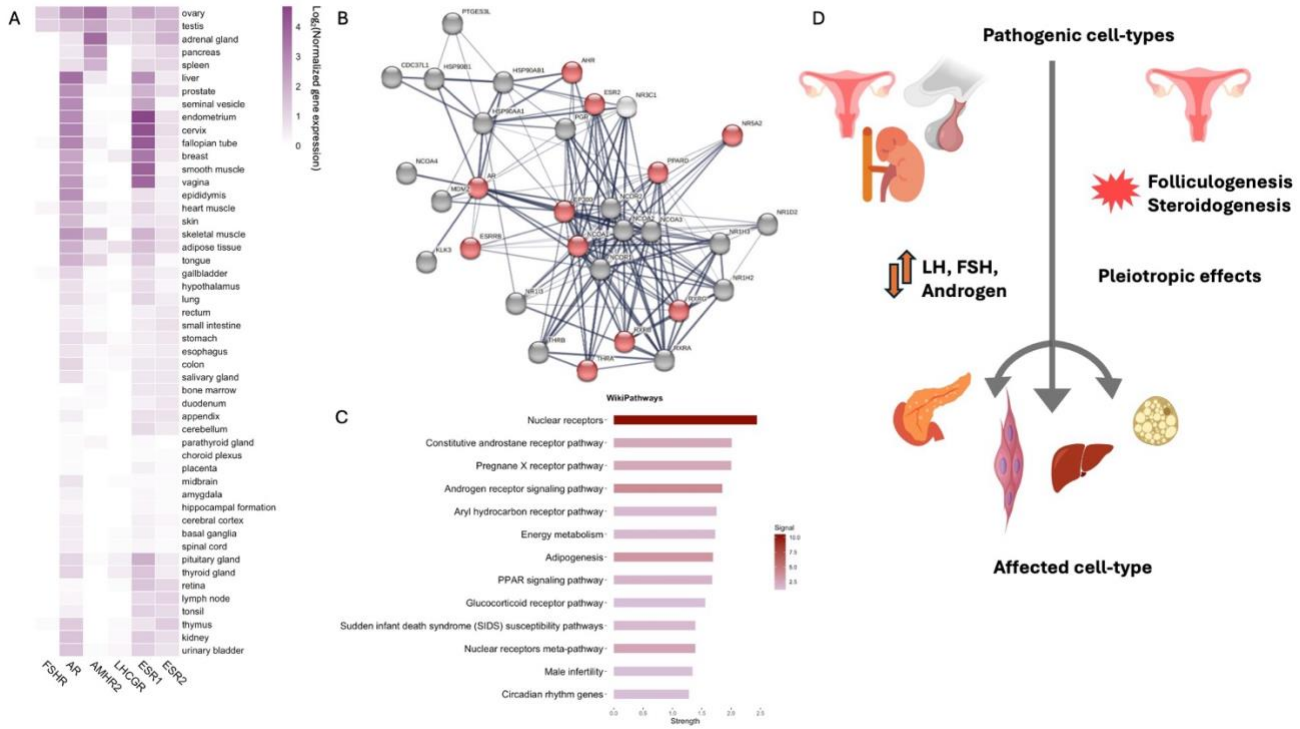

Figure 3: (A) Gene expression profiles of hormone receptors across cell-types obtained from the Human Protein Atlas (PubMed: 25613900), (B) Protein-protein interaction networks of TFs (in red) whose binding sites are enriched among pcosSNVs, (C) Pathways enriched in the interaction network. Nodes in red represent TFs whose binding sites are enriched in pcosSNVs. Signal is defined as a weighted harmonic mean between the observed/expected ratio and -log(FDR) and strength is Log10(observed / expected), (D) Proposed hormone mediated mechanism linking pathogenic cell-types in PCOS disease etiology.

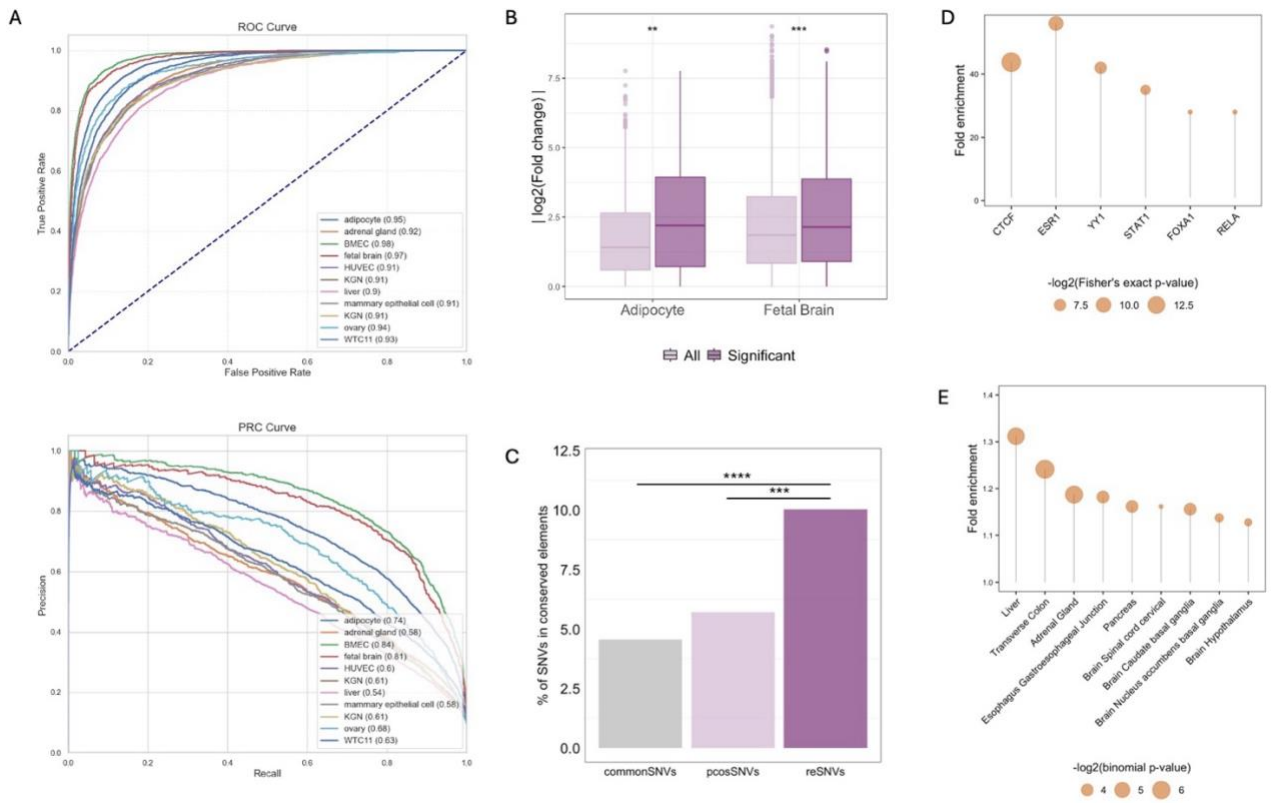

Figure 4: (A) ROC and PRC curves of eleven cell-type specific TREDNet models, (B) A comparison of fold change (alternate / reference allele) in TREDNet scores between all variants and those exhibiting significant change in enhancer activity in MPRA, using Wilcoxon test (ns: p > 0.05, *: p <= 0.05,**: p <= 0.01,***: p <= 0.001), (C) Fraction of SNVs overlapping with phastCons elements conserved across 30 primates, (D) TFs enriched among reSNVs exhibiting allele specific TF binding compared with pcosSNVs, (E) Fold enrichment of eVariants linked to reSNVs compared with pcosSNVs in GTEx tissues (p-value < 0.01).

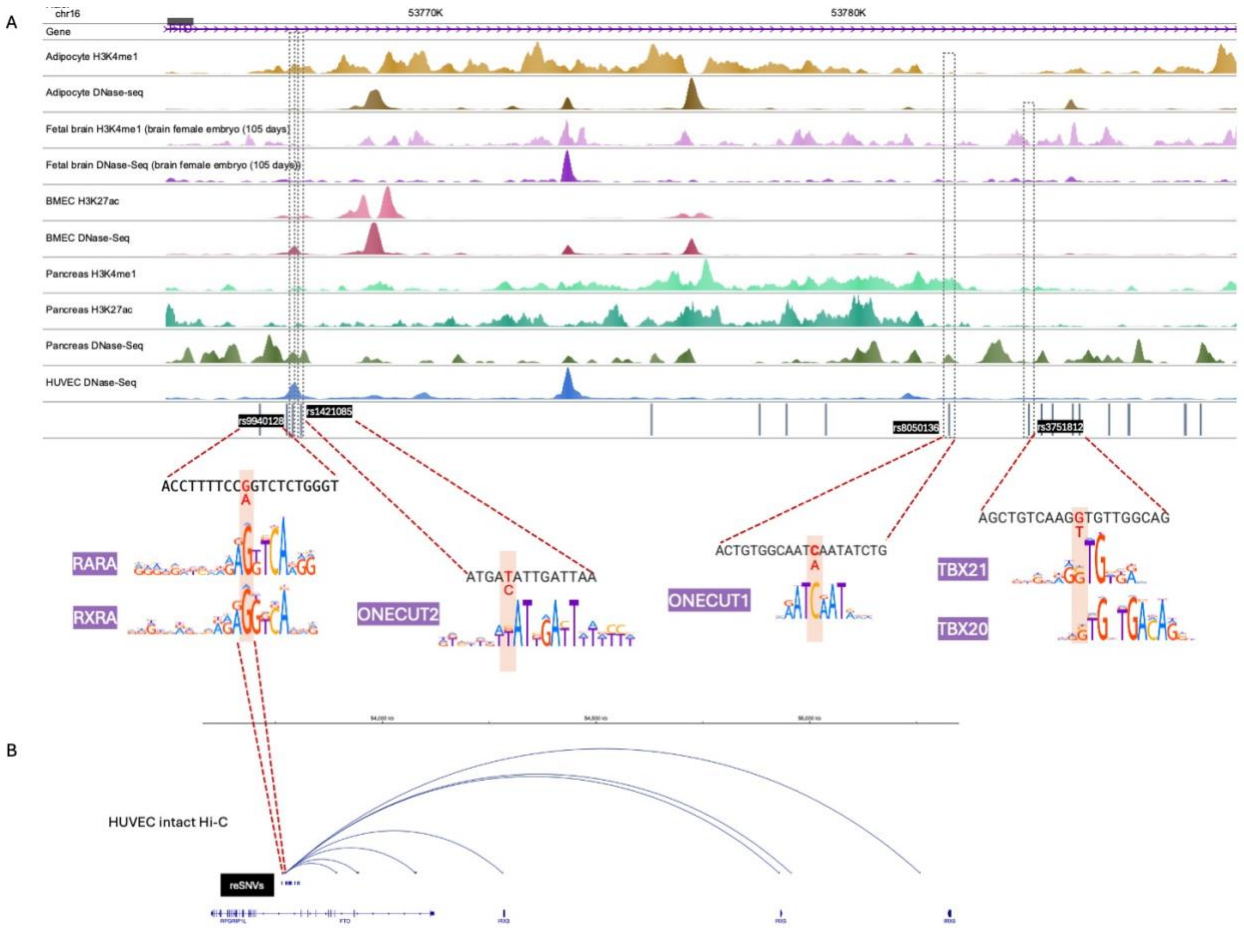

Figure 5: reSNVs in FTO locus exhibiting significant fold change in TREDNet predicted enhancer activity. (A) Overlap of reSNVs with active regulatory regions of pathogenic cell-types (B) Intact Hi-C map of chromatin interactions from reSNVs in FTO locus in HUVEC (doi:10.17989/ENCSR788FBI)