

1 **The ‘Obesity Knowledge Portal’:**
2 **an AI-powered integrative platform for exploration of obesity genetics**

3
4 Yifan Liang¹, Shiqi Ao¹, Wei Xu², Baoguo Li³, Jun-feng Bi¹, Tong-jin Zhao¹, Peng
5 Li^{3,4}, John R. Speakman^{5,6,7,8,9,*}, Guanlin Wang^{1,*}

6 **Affiliations**

7 ¹ Shanghai Key Laboratory of Metabolic Remodelling and Health, Institute of
8 Metabolism and Integrative Biology, School of Basic Medical Sciences, Centre for
9 Evolutionary Biology, Fudan University, Shanghai, 200438, China

10 ² Yunshu AI Drug Discovery Ltd, Shanghai, 201400, China

11 ³ State Key Laboratory of Metabolic Dysregulation & Prevention and Treatment of
12 Esophageal Cancer, Tianjian Laboratory of Advanced Biomedical Sciences, Academy
13 of Medical Sciences, Zhengzhou University, Zhengzhou, 450001, China

14 ⁴ Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua
15 University, Beijing, 100871, China

16 ⁵ Shenzhen Key Laboratory of Metabolic Health, Center for Energy Metabolism and
17 Reproduction, Shenzhen Institute of Advanced Technology, Chinese Academy of
18 Sciences, Shenzhen, 518055, China

19 ⁶ State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and
20 Developmental Biology, Chinese Academy of Sciences, Beijing, 100101, China

21 ⁷ Institute of Health Sciences, China Medical University, Shenyang, Liaoning, 110122,
22 China

23 ⁸ Faculty of Pharmacy, Shenzhen University of Advanced Technology, Shenzhen,
24 518107, Guangdong, China

25 ⁹ School of Biological Sciences, University of Aberdeen, AB24 2TZ, Aberdeen, UK

26

27

28 *Corresponding authors: **GW** (guanlin_wang@fudan.edu.cn)

29

30 **Summary**

31 Obesity is a chronic, multifactorial disease in which genetics plays a major role.
32 Genome-wide association studies (GWAS) have identified thousands of loci associated
33 with obesity and related traits, however, translating these findings into mechanistic
34 understanding and clinical application remains challenging. Here we present an open-
35 access and AI-powered web platform, Obesity Knowledge Portal (OKP,
36 <https://obesityknowledge.org>), which covers 166,582 curated variant-trait associations
37 mapped to 11,298 unique genes from published GWAS and UK Biobank summary
38 statistics, a manually curated database of 49 clinical-stage obesity therapeutics with
39 gene-level GWAS evidence annotation, and a domain-specific AI assistant built on
40 Retrieval-Augmented Generation (RAG) of 10,198 peer-reviewed publications. The
41 portal provides three core analytical capabilities: interactive exploration of gene- and
42 variant-level associations across multiple datasets with cross-linking to clinical-stage
43 drug targets; a genomic target prioritisation algorithm that combines statistical
44 significance and variant density to prioritise novel candidates from genes not targeted
45 by existing therapeutics, and literature-grounded interpretation of genetic and
46 pharmacological queries through a scope-restricted LLM-driven contextualisation
47 system. We demonstrate the portal's utility through case studies at the *FTO* locus and
48 an integrated pharmacogenomic analysis of the GLP-1/incretin axis. The OKP is freely
49 available and aims to support the translation of genetic discoveries into therapeutic
50 hypotheses for obesity.

51

52 **Keywords:** obesity, knowledge portal, AI, large language model, GWAS, precision
53 medicine

54

55 Introduction

56 Obesity affects over 890 million adults worldwide and contributes to the global burden
57 of type 2 diabetes, cardiovascular disease, and certain cancers^{1,2}. Twin and family
58 studies have consistently demonstrated that genetic factors account for 40-70% of the
59 variation in obesity³⁻⁵, establishing human genetics as a foundation for discovering
60 novel therapeutic targets and enabling precision medicine strategies. Since the
61 landmark identification of the *FTO* locus variant rs9939609 in 2007⁶, large-scale
62 genome-wide association studies (GWAS), including the GIANT consortium and Pan-
63 UK Biobank, have identified thousands of independent loci associated with body mass
64 index (BMI) and obesity-related traits⁷⁻¹². While these findings have suggested the
65 central role of hypothalamic and neuronal pathways in appetite regulation in
66 obesity^{7,13,14}, other studies have emphasized the broad tissue expression of these target
67 genes¹⁵⁻¹⁷, and the majority of GWAS loci have not been linked to specific biological
68 mechanisms. Very few have been translated into therapeutic targets, and these have
69 been restricted to treatments for monogenic obesities such as mutations of the *POMC*
70 gene, *PCSK1* gene and leptin receptor gene that interfere with signalling to the MC4
71 receptor. Patients with these mutations can be treated with setmelanotide which is a
72 direct agonist of the *MC4R*¹⁸.

73

74 This translational gap is particularly striking given the recent clinical success of drug
75 targets subsequently shown to have genetic support. Semaglutide, a GLP-1 receptor
76 agonist, achieves approximately 16% body weight reduction in randomised controlled
77 trials^{19,20}, Tirzepatide, a dual GLP-1/GIP receptor agonist, approximately 23%²¹ and
78 the triple agonist Retatrutide demonstrated weight loss exceeding 29% in phase 3
79 studies^{22,23}. The incretin-based therapies targeting *GLP1R*, *GIPR* and *GCGR* were not
80 initially identified as genes associated with obesity through human genetics. They
81 emerged from use as treatment for T2D, where weight loss was noted as a side effect.
82 Genetic evidence later showed that variations in or near these genes have since been
83 identified in GWAS of BMI and metabolic traits^{7,12,24,25}, and recent pharmacogenomic

84 analyses have linked such variation to interindividual differences in weight-loss
85 response^{24,26,27}. Although retrospective, this genetic corroboration nonetheless supports
86 the principle that human genetic evidence can validate and prioritise therapeutic targets.
87 This suggests that systematic mining of GWAS data for druggable genes may accelerate
88 the next generation of obesity therapeutics, however, there is no existing resource
89 providing the computational infrastructure for this purpose.

90

91 While several resources contribute to the interpretation of obesity GWAS data, none
92 provides a unified framework spanning genetic association, functional context and
93 therapeutic translation. The NHGRI-EBI GWAS Catalog²⁸ provides a comprehensive
94 repository of published associations but does not link them to drug targets or therapeutic
95 pipelines. The Type 2 Diabetes Knowledge Portal²⁹ demonstrates the value of disease-
96 focused genetic resources but does not cover the obesity pharmacological landscape.
97 Open Targets³⁰ integrates broadly across multiple diseases but lacks the mechanism of
98 action, clinical efficacy, and GWAS evidence strength per compound that are essential
99 for obesity-focused target assessment. Meanwhile, research publications on obesity
100 genetics and pharmacotherapy have grown at an extraordinary rate, with thousands of
101 papers published annually on GWAS findings, incretin pharmacology, adipose biology,
102 and mechanistic studies on obesity-associated genes, making manual synthesis by
103 individual researchers increasingly difficult. The recent emergence of large language
104 models (LLMs) as scientific tools presents an unprecedented opportunity to address this
105 challenge, enabling the transformation of static genomic annotations into dynamic,
106 context-aware interpretive summaries³¹, provided that such models can be grounded in
107 curated, domain-specific evidence rather than generic training data.

108

109 To address these challenges, we developed the Obesity Knowledge Portal (OKP,
110 <https://obesityknowledge.org>), a freely available open-access and monthly updated web
111 platform that provides an integrated environment for obesity genetics exploration and
112 drug target discovery powered by AI. The portal consolidates genome-wide association

113 data from multiple sources into a unified browser with cross-dataset comparison,
114 regional association plotting, and protein interaction network visualisation. It
115 incorporates a curated pharmacological layer linking genetic evidence to the current
116 obesity therapeutic pipeline, supported by a scoring framework for systematic target
117 prioritisation. To enable scalable evidence synthesis across the rapidly expanding
118 obesity literature, the platform deploys a Retrieval-Augmented Generation (RAG)
119 system that grounds LLM outputs in over 10,000 domain-specific publications,
120 delivering context-aware interpretation on demand. Together, these capabilities
121 establish the OKP as a unified resource for navigating from genetic association to
122 therapeutic hypothesis in obesity genetics and precision pharmacology.

123

124

125 **Results**

126 **Overview of the Obesity Knowledge Portal**

127 The Obesity Knowledge Portal integrates three complementary data layers within a
128 unified web interface (**Fig. 1A**). The genetic layer covers obesity-related variant-trait
129 associations from three primary sources. From the NHGRI-EBI GWAS Catalog, we
130 extracted all associations for obesity-related traits using keyword-based filtering across
131 the mapped traits (**Methods**), retaining variants with valid gene annotations and
132 yielding 63,960 variant-trait associations mapped to 11,298 genes across 661 distinct
133 traits, drawn from 1,145 studies reported in 366 publications. These traits span primary
134 obesity-related phenotypes, including BMI, BMI-adjusted waist-hip ratio (WHR),
135 BMI-adjusted hip circumference, BMI-adjusted waist circumference, body weight as
136 well as adiposity-associated traits including body fat percentage, visceral adipose tissue
137 mass, lean body mass, birth weight, and obesity-related metabolic traits (**Supplemental**
138 **Table 1**). The second resource comprised UK Biobank summary statistics from the
139 Neale Lab Round 2 analysis, contributing 51,998 genome-wide significant variants for
140 BMI (Field 21001) and 50,624 for BMI by bioelectrical impedance (Field 23104)³². In

141 total, the portal indexes 166,582 variant-trait associations across the three contributing
142 datasets.

143

144 The pharmacological layer comprises a manually curated database of 49 clinical-stage
145 obesity therapeutics (**Table 1**). These include six FDA-approved agents (Semaglutide,
146 Tirzepatide, Orforglipron, Liraglutide, Setmelanotide and Metreleptin), three
147 regionally approved agents in China (Mazdutide, Ecnoglutide and Beinaglutide), one
148 compound under FDA review (CagriSema), thirteen phase 1, thirteen phase 2, eleven
149 phase 3 candidates, and two withdrawn drugs (Rimonabant and Lorcaserin). The
150 pipeline spans multiple therapeutic modalities including GLP-1 receptor mono-agonists,
151 dual GLP-1/GIP agonists (Tirzepatide, MariTide, Enicepatide, VK2735, Olatorepatide,
152 THDBH120), dual GLP-1/glucagon agonists (Survodutide, Mazdutide, Pemvidutide),
153 the triple GLP-1/GIP/glucagon agonist Retatrutide, amylin-based therapies
154 (Zenagamtide, Eloralintide, Cagrilintide, Petrelintide), oral non-peptide GLP-1
155 agonists (Orforglipron, Aleniglipron), myostatin/activin pathway inhibitors
156 (Taldefgrobep alfa, Bimagrumab), a peripherally restricted CB1 inverse agonist
157 (Monlunabant) and a centrally acting CB1 inverse agonist (Rimonabant; withdrawn).
158 Each entry is annotated with the molecular target, mechanism of action, route of
159 administration, development stage, pivotal trial identifier, maximum reported weight-
160 loss efficacy, and strength of supporting GWAS evidence, with external links to
161 DrugBank³³ and ChEMBL³⁴ where available.

162

163 The knowledge layer consists of a Retrieval-Augmented Generation (RAG) system
164 built from 10,198 peer-reviewed publications relevant to obesity genetics, incretin
165 pharmacology, adipose biology, and pharmacogenomics, segmented into 150,943 text
166 segments and embedded as 384-dimensional vectors using the all-MiniLM-L6-v2³⁵
167 sentence transformer, indexed with FAISS for sub-second similarity search. This
168 architecture enables the portal's AI query system to ground its outputs in peer-reviewed,
169 domain-specific evidence rather than relying on the general training data of LLMs

170 which have a high hallucination rate.

171

172 The portal provides three primary functional modules: (i) an interactive GWAS browser
173 supporting gene and variant search, regional visualisation, and AI-powered report
174 generation; (ii) a pharmacogenomic drug discovery module integrating a curated
175 therapeutic pipeline with genomic target prioritisation and a filtered GPCR target list;
176 and (iii) an AI research assistant providing literature-grounded, scope-restricted
177 interpretation of genetic and pharmacological queries through RAG (**Fig. 1B**).

178

179 **Gene/Variant level Browser for genetic association exploration**

180 The Gene/Variant Browser provides an interactive interface for exploring genetic
181 associations across three datasets. Users can select from GWAS (default), UK Biobank
182 BMI, UK Biobank BMI Impedance and further apply trait-level filters to focus on
183 specific obesity-related phenotypes.

184

185 Searching by gene name or variant rsID provides access to multiple complementary
186 analytical views (**Supplemental Fig. 1A**): a sortable, filterable association table with
187 downloadable CSV export and direct links to PubMed records and GWAS Catalog
188 study accessions; Manhattan plot and *LocusZoom* regional association plots displaying
189 variant distribution across the chromosome and enabling fine-resolution inspection of
190 local linkage disequilibrium (LD) architecture³⁶; protein-protein interaction networks
191 from STRING³⁷ with external links to Reactome³⁸, KEGG³⁹, and GeneCards⁴⁰ for
192 pathway-level contextualisation; the AI Insight tab generates a literature-grounded
193 summary for any selected gene, synthesising biological function, therapeutic relevance
194 and key findings from the indexed corpus. The Publication Report tab compiles the
195 gene's variants, AI summary, and external resource links into a formatted PDF
196 document for offline use or sharing.

197

198 We illustrate the browser using the *FTO* gene, one of the most extensively characterised
199 obesity susceptibility genes (**Fig. 2A**). Querying *FTO* under the default GWAS returns
200 500 unique genome-wide significant variants on chromosome 16, spanning positions
201 53,750,466 - 54,117,066 and encompassing multiple distinct obesity-related traits. The
202 association table enables sorting by *P* value, chromosomal position, or effect size and
203 the full result set is downloadable for downstream analysis. The *LocusZoom* regional
204 plot provides fine-resolution LD colouring for the *FTO* locus, enabling researchers to
205 distinguish independent signals from those attributable to LD structure (**Fig. 2B**). The
206 STRING interaction network contextualises *FTO* within its functional neighbourhood,
207 revealing its central role in the m⁶A RNA methylation complex⁴¹ through interactions
208 with writers (METTL3, METTL14)⁴², erasers (ALKBH5), and readers (YTHDF1,
209 YTHDF2, IGF2BP2)⁴³, as well as connections to the established obesity-associated
210 genes *MC4R* and *TMEM18* (**Fig. 2C**). When a queried gene is a known drug target, a
211 cross-linking banner appears below the gene header, displaying all clinical-stage
212 compounds targeting that gene. For example, selecting *GLPIR* reveals 30 clinical-stage
213 compounds in the curated database that act on this target, including Semaglutide,
214 Tirzepatide, Orforglipron, and Retatrutide, providing immediate pharmacological
215 context alongside the genetic evidence (**Fig. 2D**). This bidirectional link between
216 genetic and pharmacological evidence allows researchers to move directly from a
217 GWAS signal to the corresponding therapeutic landscape. Querying *FTO* in the AI
218 Insight tab produces a literature-grounded summary covering its role as a 2-
219 oxoglutarate-dependent nucleic acid demethylase, GWAS association profile across
220 multiple adiposity traits, molecular mechanisms linking *FTO* to appetite regulation and
221 energy homeostasis and therapeutic potential including the identification of entacapone
222 as a chemical inhibitor acting through FOXO1 pathways, based on retrieved source
223 publications with PubMed links (**Supplemental Fig. 1B**).

224

225 **Drug target database and pharmacogenomic analysis**

226 The Drug Discovery module provides a comprehensive analytical interface for the
227 obesity therapeutic pipeline. The Pipeline Overview ranks reported compounds by
228 maximum reported weight-loss efficacy, coloured by development stage (**Fig. 3A**). The
229 Drug Profiles tab provides a filterable database with detailed compound information
230 including mechanism of action, administration route, pivotal trial results and external
231 links to DrugBank³³ and ChEMBL³⁴ entries, alongside expandable profiles
232 summarising key clinical evidence for each compound (**Supplemental Fig. 2A**).

233

234 To demonstrate the portal's pharmacogenomic capabilities, we examined the GLP-
235 1/incretin therapeutic axis, which currently dominates the obesity drug market^{44,45}. The
236 majority of the 49 curated compounds target the GLP-1 receptor, either alone
237 (Semaglutide, Liraglutide, Orforglipron) or in combination with the GIP receptor
238 (Tirzepatide, MariTide) or the glucagon receptor (Retatrutide, Survodutide, Mazdutide).
239 Querying these targets reveals a convergence of genetic and pharmacological evidence:
240 *GIPR* harbours multiple genome-wide significant variants representing one of the
241 strongest genetic signals among current drug targets (**Supplemental Fig. 2B**), while
242 *GLPIR* carries 10 genome-wide significant variants and both show consistent
243 association signals across multiple adiposity traits (**Fig. 2D**). The Mechanisms & AI
244 tab further demonstrates these findings through an interactive Sankey diagram mapping
245 the flow from drug to target receptor to target tissue to physiological effect, illustrating
246 how incretin-based compounds converge on hypothalamic and hindbrain circuits to
247 suppress appetite and slow gastric emptying through partially overlapping but
248 pharmacologically distinct receptor pathways (**Fig. 3B**). This integrated genetic,
249 pharmacological, and mechanistic view enables researchers to assess the depth of
250 evidence supporting each therapeutic modality within the incretin axis.

251

252 **Genomic target prioritisation engine**

253 Beyond exploring individual genes, the portal provides a systematic framework for

254 novel target discovery through the genomic target prioritisation (GTP) method
255 (**Supplemental Fig. 3A**). The GTP score combines two orthogonal components: a
256 statistical significance score $S(g)$ derived from the strongest association observed
257 across all variants mapped to gene (g) and a variant enrichment score $V(g)$ reflecting
258 the total number of genome-wide significant variants annotated to that gene into a single
259 composite metric scaled from 0 to 100 (**Methods**). To focus discovery on candidates
260 without existing pharmacological tools, the algorithm excludes 12 genes that are
261 already targets of clinical-stage obesity therapeutics in our curated database and filters
262 out multi-gene blocks and intergenic variants to ensure single-gene specificity
263 (**Methods**).

264

265 The resulting target landscape (**Supplemental Fig. 3B**) plots all scored novel candidate
266 genes by variant density against maximum statistical significance, with bubble size and
267 colour proportional to the GTP score. *FTO* ranks top overall, followed by *RFLNA* and
268 *COBLL1*. Several genes with established biological relevance to obesity and
269 metabolism appear within the top-ranked candidates. *TCF7L2*, encoding a transcription
270 factor central to pancreatic β -cell function, is one of the strongest type 2 diabetes
271 susceptibility genes⁴⁶⁻⁴⁸. *ADCY3*, encoding adenylyl cyclase 3, harbours both common-
272 variant GWAS signals and rare loss-of-function variants linked to monogenic
273 obesity^{49,50}. *TBX15* is a transcription factor regulating adipose tissue development and
274 depot-specific fat distribution⁵¹. *FAIM2* is a neural gene previously implicated in
275 appetite regulation through GWAS of childhood obesity^{52,53}. The recovery of these
276 biologically validated genes demonstrates the algorithm's capacity to prioritise
277 meaningful therapeutic candidates from genetic evidence alone. For any prioritised
278 gene, the portal generates an AI viability report summarising biological function,
279 druggability assessment, tissue expression, existing tool compounds and key challenges
280 for drug development, grounded in evidence retrieved from the indexed literature
281 (**Supplemental Fig. 3C**).

282

283 Given that GPCRs represent the most druggable and therapeutically successful protein
284 class in obesity pharmacology, a dedicated GWAS GPCRs tab cross-references
285 obesity-associated genes against the International Union of Basic and Clinical
286 Pharmacology (IUPHAR) Guide to Pharmacology curated list of 357 human GPCRs,
287 identifying those reaching genome-wide significance for obesity traits⁵⁴. This analysis
288 reveals 55 GPCRs with genome-wide significant associations and a further 8 at
289 suggestive significance ($P < 5 \times 10^{-5}$). These genes include established drug targets
290 such as *GIPR*, *MC4R*, and *CALCR* alongside orphan receptors such as *GPR61*, *GPR139*,
291 and *GHSR* (**Supplemental Table 2**). Among the GPCRs reaching genome-wide
292 significance, *GPR61* ($P = 2 \times 10^{-53}$) and *GPR139* ($P = 3 \times 10^{-26}$) stand out as orphan
293 receptors with emerging evidence of impacts on energy homeostasis but currently
294 lacking clinical-stage compounds⁵⁵⁻⁵⁷.

295

296 **AI-powered research assistant**

297 The portal incorporates a Retrieval-Augmented Generation (RAG) system to provide
298 literature-grounded interpretation of obesity genetics and pharmacology questions (**Fig.**
299 **4A**). When users submit queries through any of the portal's AI interfaces, including the
300 gene insight tab in the Gene/Variant Browser, the drug analysis module in Drug
301 Discovery, the viability report function in the Novel Target Engine and the dedicated
302 AI Assistant page, the system encodes the query using the all-MiniLM-L6-v2 sentence
303 transformer, retrieves the most relevant text segments from a FAISS index built from
304 150,943 segments across 10,198 publications, and provides these passages alongside
305 any structured portal data as context to a large language model (Claude Sonnet 4.6,
306 Anthropic), which generates a response grounded in the retrieved evidence. For
307 example, beyond genetic associations identified at the *FTO* locus, users can further
308 investigate how these effects are mediated through downstream genes, such as *IRX3*
309 and *IRX5*, using the AI assistant. This enables deeper mechanistic insights into gene
310 regulation and disease pathogenesis (**Supplemental Fig. 4A**).

311

|

312 Each response is accompanied by a list of source publications retrieved from the
313 indexed corpus, with direct links to PubMed Central, allowing users to trace claims
314 back to the underlying literature.

315

316 A key feature of the AI-powered assistant is its domain-aware query understanding.
317 Before any retrieval is performed, incoming questions are evaluated against the portal's
318 scientific scope using a two-stage filter. A keyword-based classifier first screens for
319 obesity-relevant terminology such as gene symbols, drug names, metabolic phenotypes,
320 and genetic concepts, unrecognised queries are then indexed against the RAG corpus
321 with a high similarity threshold to detect in-scope questions. Queries that are outside
322 the portal's scope, such as requests about weather, travel, general coding, or unrelated
323 medical conditions, are declined with a redirection message rather than forced through
324 the retrieval pipeline (**Fig. 4B**). This design prevents the language model from
325 fabricating answers grounded in spurious context, a known failure mode of general-
326 purpose chatbots when confronted with out-of-domain prompts. For in-scope queries,
327 the system applies adaptive retrieval strategies to handle short or ambiguous inputs.
328 When a query is too brief to provide sufficient semantic context, for example, a simple
329 gene symbol such as *FTO*, the system automatically generates several expanded
330 variants by appending domain terms such as '*FTO* obesity genetics' or '*FTO* BMI
331 GWAS', retrieves passages for each variant, and returns results from the highest-
332 scoring version. Together, these mechanisms yield an AI research assistant that behaves
333 as a focused domain expert, restricted in scope, grounded in retrieved evidence and
334 transparent about its sources rather than an unconstrained conversational agent.

335

336 **Discussion**

337 Human genetics has transformed our understanding of obesity, however, the translation
338 of this evidence into therapeutic discovery has been restricted to specific drugs targeting
339 monogenic issues – for example, Setmelanotide, an *MC4R* agonist used to treat patients

340 with rare biallelic mutations in *POMC*⁵⁸. Since the identification of *FTO* in 2007⁶,
341 GWAS have reported thousands of loci associated with BMI and related obesity traits.
342 In parallel, the clinical success of incretin-based drugs has demonstrated that targets
343 supported by human genetic evidence can achieve substantial weight loss⁵⁹. Despite the
344 link between genetic discovery and pharmacological outcome, methods to bridge these
345 remain limited. GWAS catalogues provide variant-level summaries but lack
346 pharmacological context, drug target databases such as DrugBank³³ and Open
347 Targets^{30,60} contain extensive pharmacological annotations but are not ideal for the
348 genetic architecture of obesity, and clinical trial registries capture development stage
349 without linking to the underlying human genetics. The Obesity Knowledge Portal
350 addresses this clear gap providing integrated access to 166,582 variant-trait associations
351 with a manually curated database of 49 clinical-stage obesity compounds, a genomic
352 target prioritisation method, and a literature-based AI assistant indexed 10,198
353 publications.

354

355 Notably, of the 49 clinical-stage compounds curated in the portal, the majority act on
356 the GLP-1/incretin axis, either as monotherapies directed at the GLP-1 receptor
357 (Semaglutide, Liraglutide, Orforglipron), as dual agonists co-targeting the GIP receptor
358 (Tirzepatide, MariTide) or the glucagon receptor (Survodutide, Mazdutide), or as a
359 triple GLP-1/GIP/glucagon agonist (Retatrutide). The remaining compounds are
360 distributed across *MC4R*, leptin, endocannabinoid, amylin, myostatin and *GPR119*
361 pathways⁴⁵. Incretin-based therapies have shown strong efficacy in clinical trials, with
362 semaglutide achieving approximately 16% mean weight reduction^{19,20}, Tirzepatide
363 23%²¹ and Retatrutide 29%²². However, it is well-known that human genetic evidence
364 base is considerably broader than this incretin axis reflects and our gene prioritisation
365 algorithms also identify multiple high-scoring genes with established roles in energy
366 homeostasis which are not currently targeted by any clinical-stage compound. This
367 contrast between the breadth of genetic evidence and the narrowness of current
368 therapeutic development highlights the value of integrative resources that can

|

369 systematically connect GWAS findings to pharmacological opportunities beyond the
370 incretin axis.

371

372 The genomic target prioritisation analysis in the portal further provides a framework
373 for examining how current therapeutic development aligns with the broader landscape
374 of human genetic evidence for obesity. Several of the highest-scoring candidates are
375 well-characterised genes associated with metabolic traits for which no reported clinical-
376 stage obesity therapeutic is currently in development. *TCF7L2*, the strongest locus for
377 T2D and a central regulator of pancreatic β -cell function⁴⁶⁻⁴⁸, is not currently under an
378 obesity-directed development programme despite its strong and reproducible GWAS
379 signal. *ADCY3* is of particular interest, as it harbours both common-variant GWAS
380 signals and rare loss-of-function alleles linked to monogenic obesity^{49,50}, indicating a
381 convergence of evidence that supports hypothalamic cAMP signalling⁵⁷ as a
382 mechanistically potential axis for therapeutic exploration. The high ranking of *TBX15*,
383 a transcription factor regulating adipocyte development and depot-specific fat
384 distribution⁵¹, suggests adipose tissue requires more pharmacological attention apart
385 from the predominantly central nervous system-focused therapeutic development.
386 *FAIM2*, implicated in appetite regulation through GWAS of early-onset obesity,
387 suggests the involvement of neural circuits distinct from the hypothalamic melanocortin
388 and incretin pathways⁵³. 34% of the FDA-approved drugs target GPCRs^{61,62}, applying
389 the GPCR filter prioritised *GPR61* and *GPR139*, two orphan receptors with genome-
390 wide significant associations, which exhibits a strong genome-wide significant
391 association signal are not being evaluated currently in clinical-stage obesity trials. More
392 recently, multi-target combinations also suggest the success of ‘incretin +’, genes with
393 strong genetic contribution may also facilitate this multiple targets combination⁶³.
394 Taken together, these results suggest that the portal can help researchers to evaluate the
395 strength of investigation, and guide the translation of these findings towards
396 pharmacological evaluation and drug development.

397

|

398 At a broader level, the AI assistant integrated into the portal exemplifies how large
399 language models can be adapted responsibly to support the understanding of obesity
400 genetics and pathology. By constraining generation to a curated corpus of obesity
401 literature, applying a domain-specific scope filter, and returning every retrieved source
402 alongside its response, the assistant provides a form of literature synthesis that is
403 transparent, traceable, and grounded in evidence that users can easily independently
404 verify. This architecture offers a template for similar domain-focused scientific
405 assistants in other disease areas^{64,65}. Positioning such an assistant within an integrated
406 genetic and pharmacological environment, rather than as a standalone conversational
407 interface, further enhances its utility by allowing researchers to move fluidly between
408 structured data analysis and narrative literature interpretation.

409

410 By bringing together human genetic evidence, pharmacological context, and literature-
411 grounded analysis within a single integrated environment, the Obesity Knowledge
412 Portal is intended to serve as a shared human genetics and pharmacotherapy resource
413 for the research community working on metabolism and obesity. This portal will
414 support hypothesis generation, accelerate target evaluation, and facilitate
415 communication between the genetics, pharmacology, and clinical research
416 communities, thereby contributing to the broader translational effort of converting
417 genetic discovery into effective therapeutic intervention for obesity and its associated
418 metabolic diseases.

419

420

421 **Methods**

422 **Data sources and integration**

423 The portal integrates three primary genetic datasets covering obesity and adiposity traits.
424 GWAS Catalog full association data (version 1.2, April 2026) were downloaded from
425 the NHGRI-EBI GWAS Catalog and filtered to retain variant-trait associations relevant

|

426 to obesity, BMI, body fat distribution, and related adiposity phenotypes. Variants
427 without valid gene annotations were removed, yielding 63,960 variant-trait associations
428 mapped to 11,298 genes across 661 traits and 1,145 studies. UK Biobank summary
429 statistics were obtained from the Neale Lab Round 2 release for Field 21001 (body mass
430 index; 51,998 genome-wide significant variants) and Field 23104 (BMI by impedance;
431 50,624 variants)³². The combined dataset comprises 166,582 variant-trait associations.

432

433 **Drug target database**

434 The drug target database was manually curated from primary literature,
435 ClinicalTrials.gov and U.S. Food and Drug Administration (FDA) records, comprising
436 49 clinical-stage compounds on 11 unique molecular targets. Each entry includes drug
437 name, molecular target, mechanism of action, pharmacological class, administration
438 route, development stage, sponsoring company, pivotal clinical trial, maximum
439 reported weight loss efficacy, and external database identifiers (DrugBank³³ and
440 ChEMBL³⁴) if available. Clinical efficacy values correspond to the maximum reported
441 mean percentage body weight reduction at the highest evaluated dose from the
442 referenced clinical trial for each compound. For compounds in earlier development
443 stages, values reflect the best available published data and may derive from shorter-
444 duration or dose-finding studies. Compounds approved for non-obesity indications
445 (metreleptin) or without reported weight-loss endpoints (taldefgrobep alfa, macupatide)
446 are not listed in the barchart, but kept in the drug profile database. Entries are manually
447 updated as new clinical data become available.

448

449 **Genomic target prioritisation (GTP) score**

450 The GTP score integrates two normalised components of human genetic evidence:
451 statistical significance and variant density. For each candidate gene g , $P_{min}(g)$
452 denotes the smallest P value observed across all variants mapped to g , and $N(g)$
453 denotes the total number of variants mapped to g that reach genome-wide significance.

|

454 A capped log-transformed significance score is computed as

$$455 \quad S(g) = \min(-\log_{10} P_{min}(g), 100) / \max_g \{ \min(-\log_{10} P_{min}(g), 100) \}$$

456 where the cap of 100 limits the influence of extreme outliers.

457 A log-transformed variant density score is computed as

$$458 \quad V(g) = \log_2(N(g) + 1) / \max_g \{ \log_2(N(g) + 1) \}$$

459 Both components are normalised to [0, 1] using the maximum value across all scored
460 candidate genes. The composite GTP score is the equally weighted average scaled to a
461 0–100 range:

$$462 \quad GTP_{(g)} = \frac{[S_{(g)} + V_{(g)}]}{2} * 100$$

463 We exclude 12 genes that are already targets of clinical-stage obesity therapeutics in
464 the curated drug database (*GLPIR*, *GIPR*, *GCGR*, *MC4R*, *ACVR2B*, *CNRI*, *HTR2C*,
465 *LEP*, *LEPR*, *CALCR*, *MSTN* and *IGFIR*) to focus on novel candidates. Multi-gene
466 blocks entries and intergenic variants are filtered prior to scoring to ensure single-gene
467 specificity. Equal weighting of the two components is used.

468

469 **Literature corpus and RAG index**

470 The portal incorporates a retrieval-augmented generation system indexed over 10,198
471 obesity-relevant peer-reviewed publications (465 full-text PDFs and 9,733 text-
472 extracted articles) obtained via PubMed Central. Documents were chunked into
473 overlapping text segments and embedded using the all-MiniLM-L6-v2 sentence
474 transformer (384-dimensional embeddings), producing a FAISS index of 150,943
475 vectors. At query time, the system encodes the user question using the same embedding
476 model and retrieves the top-ranked segments by cosine similarity. The top segments are
477 passed as context to the generative model (Claude Sonnet 4.6, Anthropic) along with
478 any structured portal data relevant to the query, and the model generates a response
479 grounded in the retrieved passages. All unique source documents corresponding to the

480 retrieved chunks are returned to the user interface and displayed as a references list with
481 clickable links to PubMed Central.

482

483 **Domain-aware query in the AI assistant**

484 To confine the AI assistant to obesity-relevant topics, incoming queries are evaluated
485 by a two-stage filter before retrieval is performed. A keyword-based classifier first
486 screens the query against a curated list of obesity, genetics, and pharmacology terms,
487 as well as a second list of out-of-scope phrases such as requests about weather, travel,
488 or unrelated medical conditions. Queries that fail keyword matching are then probed
489 against the RAG corpus at a high similarity threshold (cosine similarity ≥ 0.55) to detect
490 in-scope questions and queries that still fail this flow are declined with a redirection
491 message. For in-scope queries, the system applies adaptive retrieval strategies to handle
492 short and ambiguous inputs. When the original query scores below a confidence
493 threshold or consists of only a few tokens (for example, a single gene symbol) or
494 retrieves passages below a confidence threshold (cosine similarity < 0.60), the system
495 generates several expanded query variants by appending domain terms such as ‘obesity
496 genetics’, ‘BMI GWAS’, or ‘adipose metabolism’ runs each variant against the index
497 and returns the results from the highest-scoring variant.

498

499 **Portal implementation**

500 The portal is implemented as a Streamlit web application (Python 3.12.3) and deployed
501 on a DigitalOcean virtual machine running Ubuntu 24.04.4 LTS behind an nginx
502 reverse proxy with Let's Encrypt TLS termination. Interactive visualisations use Plotly
503 for Manhattan plots, bubble charts, and Sankey diagrams; *LocusZoom* plots are
504 rendered via the University of Michigan LocusZoom.js API³⁶. Protein-protein
505 interaction networks are retrieved from STRING³⁷. The GPCR target filter cross-
506 references obesity-associated genes against the IUPHAR/BPS Guide to Pharmacology
507 curated list of 357 human GPCRs⁵⁴. The portal is freely accessible at

|

508 <https://obesityknowledge.org>.

509

510 References

- 511 1. Afshin, A., Forouzanfar, M.H., Reitsma, M.B., Sur, P., Estep, K., Lee, A.,
512 Marczak, L., Mokdad, A.H., Moradi-Lakeh, M., Naghavi, M., et al. (2017).
513 Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *N*
514 *Engl J Med* 377, 13-27. 10.1056/NEJMoa1614362.
- 515 2. Phelps, N.H., Singleton, R.K., Zhou, B., Heap, R.A., Mishra, A., Bennett, J.E.,
516 Paciorek, C.J., Lhoste, V.P.F., Carrillo-Larco, R.M., Stevens, G.A., et al.
517 (2024). Worldwide trends in underweight and obesity from 1990 to 2022: a
518 pooled analysis of 3663 population-representative studies with 222 million
519 children, adolescents, and adults. *The Lancet* 403, 1027-1050.
520 10.1016/S0140-6736(23)02750-2.
- 521 3. Maes, H.H., Neale, M.C., and Eaves, L.J. (1997). Genetic and environmental
522 factors in relative body weight and human adiposity. *Behav Genet* 27, 325-
523 351. 10.1023/a:1025635913927.
- 524 4. Loos, R.J.F., and Yeo, G.S.H. (2022). The genetics of obesity: from discovery
525 to biology. *Nat Rev Genet* 23, 120-133. 10.1038/s41576-021-00414-z.
- 526 5. Silventoinen, K., Rokholm, B., Kaprio, J., and Sørensen, T.I. (2010). The
527 genetic and environmental influences on childhood obesity: a systematic
528 review of twin and adoption studies. *Int J Obes (Lond)* 34, 29-40.
529 10.1038/ijo.2009.177.
- 530 6. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M.,
531 Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W., et al.
532 (2007). A common variant in the FTO gene is associated with body mass
533 index and predisposes to childhood and adult obesity. *Science* 316, 889-894.
534 10.1126/science.1141634.
- 535 7. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R.,
536 Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al. (2015). Genetic
537 studies of body mass index yield new insights for obesity biology. *Nature* 518,
538 197-206. 10.1038/nature14177.
- 539 8. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon,
540 M.N., Frayling, T.M., Hirschhorn, J., Yang, J., and Visscher, P.M. (2018).
541 Meta-analysis of genome-wide association studies for height and body mass
542 index in ~700000 individuals of European ancestry. *Hum Mol Genet* 27,
543 3641-3649. 10.1093/hmg/ddy271.
- 544 9. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell,
545 J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of
546 genome-wide association studies for body fat distribution in 694 649
547 individuals of European ancestry. *Hum Mol Genet* 28, 166-174.
548 10.1093/hmg/ddy327.
- 549 10. Huang, J., Huffman, J.E., Huang, Y., Do Valle, Í., Assimes, T.L., Raghavan,
550 S., Voight, B.F., Liu, C., Barabási, A.L., Huang, R.D.L., et al. (2022).
551 Genomics and phenomics of body mass index reveals a complex disease
552 network. *Nat Commun* 13, 7973. 10.1038/s41467-022-35553-2.

- 553 11. Karczewski, K.J., Gupta, R., Kanai, M., Lu, W., Tsuo, K., Wang, Y., Walters,
554 R.K., Turley, P., Callier, S., Shah, N.N., et al. (2025). Pan-UK Biobank
555 genome-wide association analyses enhance discovery and resolution of
556 ancestry-enriched effects. *Nat Genet* 57, 2408-2417. 10.1038/s41588-025-
557 02335-7.
- 558 12. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G.,
559 Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Mägi, R., et al.
560 (2010). Association analyses of 249,796 individuals reveal 18 new loci
561 associated with body mass index. *Nat Genet* 42, 937-948. 10.1038/ng.686.
- 562 13. Willer, C.J., Speliotes, E.K., Loos, R.J., Li, S., Lindgren, C.M., Heid, I.M.,
563 Berndt, S.I., Elliott, A.L., Jackson, A.U., Lamina, C., et al. (2009). Six new loci
564 associated with body mass index highlight a neuronal influence on body
565 weight regulation. *Nat Genet* 41, 25-34. 10.1038/ng.287.
- 566 14. Timshel, P.N., Thompson, J.J., and Pers, T.H. (2020). Genetic mapping of
567 etiologic brain cell types for obesity. *Elife* 9. 10.7554/eLife.55851.
- 568 15. Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi,
569 P., Park, Y., Parsana, P., Segrè, A.V., et al. (2017). Genetic effects on gene
570 expression across human tissues. *Nature* 550, 204-213.
571 10.1038/nature24277.
- 572 16. Consortium, T.G., Aguet, F., Anand, S., Ardlie, K.G., Gabriel, S., Getz, G.A.,
573 Graubert, A., Hadley, K., Handsaker, R.E., Huang, K.H., et al. (2020). The
574 GTEx Consortium atlas of genetic regulatory effects across human tissues.
575 *Science* 369, 1318-1330. doi:10.1126/science.aaz1776.
- 576 17. Wu, Y., Rudolf, A.M., Wang, L., Zhang, Z., Niu, C., Xia, F., Li, M., Wang, A.,
577 Derous, D., Mitchell, S.E., et al. (2026). The susceptibility to weight gain in
578 male C57BL/6N mice fed high fat diet is due to a small set of genes
579 predominantly not expressed in the brain: I target identification. *Proceedings*
580 *of the National Academy of Sciences*, accepted.
- 581 18. Markham, A. (2021). Setmelanotide: First Approval. *Drugs* 81, 397-403.
582 10.1007/s40265-021-01470-9.
- 583 19. Wilding, J.P.H., Batterham, R.L., Calanna, S., Davies, M., Gaal, L.F.V.,
584 Lingvay, I., McGowan, B.M., Rosenstock, J., Tran, M.T.D., Wadden, T.A., et
585 al. (2021). Once-Weekly Semaglutide in Adults with Overweight or Obesity.
586 *New England Journal of Medicine* 384, 989-1002.
587 doi:10.1056/NEJMoa2032183.
- 588 20. Garvey, W.T., Batterham, R.L., Bhatta, M., Buscemi, S., Christensen, L.N.,
589 Frias, J.P., Jódar, E., Kandler, K., Rigas, G., Wadden, T.A., and Wharton, S.
590 (2022). Two-year effects of semaglutide in adults with overweight or obesity:
591 the STEP 5 trial. *Nat Med* 28, 2083-2091. 10.1038/s41591-022-02026-4.
- 592 21. Jastreboff, A.M., Aronne, L.J., Ahmad, N.N., Wharton, S., Connery, L., Alves,
593 B., Kiyosue, A., Zhang, S., Liu, B., Bunck, M.C., and Stefanski, A. (2022).
594 Tirzepatide Once Weekly for the Treatment of Obesity. *New England Journal*
595 *of Medicine* 387, 205-216. doi:10.1056/NEJMoa2206038.
- 596 22. Jastreboff, A.M., Kaplan, L.M., Frías, J.P., Wu, Q., Du, Y., Gurbuz, S.,

- 597 Coskun, T., Haupt, A., Milicevic, Z., and Hartman, M.L. (2023). Triple-
598 Hormone-Receptor Agonist Retatrutide for Obesity — A Phase 2 Trial. *New*
599 *England Journal of Medicine* 389, 514-526. doi:10.1056/NEJMoa2301972.
- 600 23. Giblin, K., Kaplan, L.M., Somers, V.K., Le Roux, C.W., Hunter, D.J., Wu, Q.,
601 Lalonde, A., Ahmad, N., and Bethel, M.A. (2026). Retatrutide for the
602 treatment of obesity, obstructive sleep apnea and knee osteoarthritis:
603 Rationale and design of the TRIUMPH registrational clinical trials. *Diabetes*
604 *Obes Metab* 28, 83-93. 10.1111/dom.70209.
- 605 24. Gao, W., Liu, L., Huh, E., Gbahou, F., Cecon, E., Oshima, M., Houzé, L.,
606 Katsonis, P., Hegron, A., Fan, Z., et al. (2023). Human GLP1R variants
607 affecting GLP1R cell surface expression are associated with impaired glucose
608 control and increased adiposity. *Nat Metab* 5, 1673-1684. 10.1038/s42255-
609 023-00889-6.
- 610 25. Lagou, V., Jiang, L., Ulrich, A., Zudina, L., González, K.S.G., Balkhiyarova,
611 Z., Faggian, A., Maina, J.G., Chen, S., Todorov, P.V., et al. (2023). GWAS of
612 random glucose in 476,326 individuals provide insights into diabetes
613 pathophysiology, complications and treatment stratification. *Nat Genet* 55,
614 1448-1461. 10.1038/s41588-023-01462-3.
- 615 26. Dawed, A.Y., Mari, A., Brown, A., McDonald, T.J., Li, L., Wang, S., Hong,
616 M.G., Sharma, S., Robertson, N.R., Mahajan, A., et al. (2023).
617 Pharmacogenomics of GLP-1 receptor agonists: a genome-wide analysis of
618 observational data and large randomised controlled trials. *Lancet Diabetes*
619 *Endocrinol* 11, 33-41. 10.1016/s2213-8587(22)00340-0.
- 620 27. Kizilkaya, H.S., Sørensen, K.V., Madsen, J.S., Lindquist, P., Douros, J.D.,
621 Bork-Jensen, J., Berghella, A., Gerlach, P.A., Gasbjerg, L.S., Mokrosiński, J.,
622 et al. (2024). Characterization of genetic variants of GIPR reveals a
623 contribution of β -arrestin to metabolic phenotypes. *Nat Metab* 6, 1268-1281.
624 10.1038/s42255-024-01061-4.
- 625 28. Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T.,
626 Güneş, O., Hall, P., Hayhurst, J., et al. (2023). The NHGRI-EBI GWAS
627 Catalog: knowledgebase and deposition resource. *Nucleic Acids Res* 51,
628 D977-d985. 10.1093/nar/gkac1010.
- 629 29. Costanzo, M.C., von Grotthuss, M., Massung, J., Jang, D., Caulkins, L.,
630 Koesterer, R., Gilbert, C., Welch, R.P., Kudtarkar, P., Hoang, Q., et al. (2023).
631 The Type 2 Diabetes Knowledge Portal: An open access genetic resource
632 dedicated to type 2 diabetes and related traits. *Cell Metab* 35, 695-710.e696.
633 10.1016/j.cmet.2023.03.001.
- 634 30. Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Baker, J., Malangone,
635 C., Lopez, I., Miranda, A., Cruz-Castillo, C., Fumis, L., et al. (2023). The next-
636 generation Open Targets Platform: reimaged, redesigned, rebuilt. *Nucleic*
637 *Acids Res* 51, D1353-d1359. 10.1093/nar/gkac1046.
- 638 31. Thirunavukarasu, A.J., Ting, D.S.J., Elangovan, K., Gutierrez, L., Tan, T.F.,
639 and Ting, D.S.W. (2023). Large language models in medicine. *Nature*
640 *Medicine* 29, 1930-1940. 10.1038/s41591-023-02448-8.

- 641 32. Lab, N. (2018). UK Biobank — Neale Lab. <http://www.nealelab.is/uk-biobank/>.
- 642 33. Knox, C., Wilson, M., Klinger, C.M., Franklin, M., Oler, E., Wilson, A., Pon, A.,
643 Cox, J., Chin, N.E.L., Strawbridge, S.A., et al. (2024). DrugBank 6.0: the
644 DrugBank Knowledgebase for 2024. *Nucleic Acids Res* 52, D1265-d1275.
645 10.1093/nar/gkad976.
- 646 34. Zdrzil, B., Felix, E., Hunter, F., Manners, E.J., Blackshaw, J., Corbett, S.,
647 de Veij, M., Ioannidis, H., Lopez, D.M., Mosquera, Juan F., et al. (2023). The
648 ChEMBL Database in 2023: a drug discovery platform spanning multiple
649 bioactivity data types and time periods. *Nucleic Acids Research* 52, D1180-
650 D1192. 10.1093/nar/gkad1004.
- 651 35. Reimers, N., and Gurevych, I. (2019). Sentence-bert: Sentence embeddings
652 using siamese bert-networks. pp. 3982-3992.
- 653 36. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt,
654 T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom:
655 regional visualization of genome-wide association scan results. *Bioinformatics*
656 26, 2336-2337. 10.1093/bioinformatics/btq419.
- 657 37. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R.,
658 Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The
659 STRING database in 2023: protein-protein association networks and
660 functional enrichment analyses for any sequenced genome of interest.
661 *Nucleic Acids Res* 51, D638-d646. 10.1093/nar/gkac1000.
- 662 38. Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro,
663 A., Griss, J., Sevilla, C., Matthews, L., Gong, C., et al. (2022). The reactome
664 pathway knowledgebase 2022. *Nucleic Acids Res* 50, D687-d692.
665 10.1093/nar/gkab1028.
- 666 39. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-
667 Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and
668 genomes. *Nucleic Acids Res* 51, D587-d592. 10.1093/nar/gkac963.
- 669 40. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich,
670 S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., et al. (2016). The GeneCards
671 Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr*
672 *Protoc Bioinformatics* 54, 1.30.31-31.30.33. 10.1002/cpbi.5.
- 673 41. Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan,
674 T., Yang, Y.G., and He, C. (2011). N6-methyladenosine in nuclear RNA is a
675 major substrate of the obesity-associated FTO. *Nat Chem Biol* 7, 885-887.
676 10.1038/nchembio.687.
- 677 42. Zhao, X., Yang, Y., Sun, B.F., Shi, Y., Yang, X., Xiao, W., Hao, Y.J., Ping,
678 X.L., Chen, Y.S., Wang, W.J., et al. (2014). FTO-dependent demethylation of
679 N6-methyladenosine regulates mRNA splicing and is required for
680 adipogenesis. *Cell Res* 24, 1403-1419. 10.1038/cr.2014.151.
- 681 43. Merkestein, M., Laber, S., McMurray, F., Andrew, D., Sachse, G., Sanderson,
682 J., Li, M., Usher, S., Sellayah, D., Ashcroft, F.M., and Cox, R.D. (2015). FTO
683 influences adipogenesis by regulating mitotic clonal expansion. *Nat Commun*
684 6, 6792. 10.1038/ncomms7792.

- 685 44. Drucker, D.J. (2022). GLP-1 physiology informs the pharmacotherapy of
686 obesity. *Mol Metab* 57, 101351. 10.1016/j.molmet.2021.101351.
- 687 45. Müller, T.D., Blüher, M., Tschöp, M.H., and DiMarchi, R.D. (2022). Anti-
688 obesity drug discovery: advances and challenges. *Nat Rev Drug Discov* 21,
689 201-223. 10.1038/s41573-021-00337-8.
- 690 46. Grant, S.F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu,
691 A., Sainz, J., Helgason, A., Stefansson, H., Emilsson, V., Helgadottir, A., et al.
692 (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of
693 type 2 diabetes. *Nat Genet* 38, 320-323. 10.1038/ng1732.
- 694 47. Florez, J.C., Jablonski, K.A., Bayley, N., Pollin, T.I., de Bakker, P.I.,
695 Shuldiner, A.R., Knowler, W.C., Nathan, D.M., and Altshuler, D. (2006).
696 TCF7L2 polymorphisms and progression to diabetes in the Diabetes
697 Prevention Program. *N Engl J Med* 355, 241-250. 10.1056/NEJMoa062418.
- 698 48. Mahajan, A., Spracklen, C.N., Zhang, W., Ng, M.C.Y., Petty, L.E., Kitajima,
699 H., Yu, G.Z., Rieger, S., Speidel, L., Kim, Y.J., et al. (2022). Multi-ancestry
700 genetic study of type 2 diabetes highlights the power of diverse populations
701 for discovery and translation. *Nat Genet* 54, 560-572. 10.1038/s41588-022-
702 01058-3.
- 703 49. Saeed, S., Bonnefond, A., Tamanini, F., Mirza, M.U., Manzoor, J., Janjua,
704 Q.M., Din, S.M., Gaitan, J., Milochau, A., Durand, E., et al. (2018). Loss-of-
705 function mutations in ADCY3 cause monogenic severe obesity. *Nat Genet* 50,
706 175-179. 10.1038/s41588-017-0023-6.
- 707 50. Grarup, N., Moltke, I., Andersen, M.K., Dalby, M., Vitting-Seerup, K., Kern, T.,
708 Mahendran, Y., Jørsboe, E., Larsen, C.V.L., Dahl-Petersen, I.K., et al. (2018).
709 Loss-of-function variants in ADCY3 increase risk of obesity and type 2
710 diabetes. *Nat Genet* 50, 172-174. 10.1038/s41588-017-0022-7.
- 711 51. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E.,
712 Mägi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., et al.
713 (2015). New genetic loci link adipose and insulin biology to body fat
714 distribution. *Nature* 518, 187-196. 10.1038/nature14132.
- 715 52. Bradfield, J.P., Taal, H.R., Timpson, N.J., Scherag, A., Lecoeur, C.,
716 Warrington, N.M., Hypponen, E., Holst, C., Valcarcel, B., Thiering, E., et al.
717 (2012). A genome-wide association meta-analysis identifies new childhood
718 obesity loci. *Nat Genet* 44, 526-531. 10.1038/ng.2247.
- 719 53. Littleton, S.H., Trang, K.B., Volpe, C.M., Cook, K., DeBruyne, N., Maguire,
720 J.A., Weidekamp, M.A., Hodge, K.M., Boehm, K., Lu, S., et al. (2024).
721 Variant-to-function analysis of the childhood obesity chr12q13 locus
722 implicates rs7132908 as a causal variant within the 3' UTR of FAIM2. *Cell*
723 *Genom* 4, 100556. 10.1016/j.xgen.2024.100556.
- 724 54. Harding, S.D., Armstrong, J.F., Faccenda, E., Southan, C., Alexander, S.P.H.,
725 Davenport, A.P., Spedding, M., and Davies, J.A. (2024). The IUPHAR/BPS
726 Guide to PHARMACOLOGY in 2024. *Nucleic Acids Res* 52, D1438-d1449.
727 10.1093/nar/gkad944.
- 728 55. Tsang, C.H., De Rosa, A., and Koziellewicz, P. (2025). Identification and

- 729 molecular characterization of missense mutations in orphan G protein-
730 coupled receptor GPR61 occurring in severe obesity. *Mol Pharmacol* 107,
731 100026. 10.1016/j.molpha.2025.100026.
- 732 56. Nogueira, P.A.S., Moura-Assis, A., Zanesco, A.M., Bombassaro, B., Gallo-
733 Ferraz, A.L., Simões, M.R., Engel, D.F., Razolli, D.S., Gaspar, J.M., Junior,
734 J.D., and Velloso, L.A. (2023). The orphan G protein-coupled receptor,
735 GPR139, is expressed in the hypothalamus and is involved in the regulation
736 of body mass, blood glucose, and insulin. *Neurosci Lett* 792, 136955.
737 10.1016/j.neulet.2022.136955.
- 738 57. Felix, J.F., Bradfield, J.P., Monnerieu, C., van der Valk, R.J., Stergiakouli, E.,
739 Chesi, A., Gaillard, R., Feenstra, B., Thiering, E., Kreiner-Møller, E., et al.
740 (2016). Genome-wide association analysis identifies three new susceptibility
741 loci for childhood body mass index. *Hum Mol Genet* 25, 389-403.
742 10.1093/hmg/ddv472.
- 743 58. Clément, K., van den Akker, E., Argente, J., Bahm, A., Chung, W.K.,
744 Connors, H., De Waele, K., Farooqi, I.S., Gonneau-Lejeune, J., Gordon, G.,
745 et al. (2020). Efficacy and safety of setmelanotide, an MC4R agonist, in
746 individuals with severe obesity due to LEPR or POMC deficiency: single-arm,
747 open-label, multicentre, phase 3 trials. *Lancet Diabetes Endocrinol* 8, 960-
748 970. 10.1016/s2213-8587(20)30364-8.
- 749 59. Bonnefond, A., Bruner, W.S., Grant, S.F., Morandi, A., and Froguel, P.
750 (2026). The genetics of obesity: aetiology, prevention and therapy. *Nature*
751 *Metabolism*, 1-17.
- 752 60. Ghousaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M.,
753 Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A., et al.
754 (2021). Open Targets Genetics: systematic identification of trait-associated
755 genes using large-scale genetics and functional genomics. *Nucleic Acids Res*
756 49, D1311-d1320. 10.1093/nar/gkaa840.
- 757 61. Lorente, J.S., Sokolov, A.V., Ferguson, G., Schiöth, H.B., Hauser, A.S., and
758 Gloriam, D.E. (2025). GPCR drug discovery: new agents, targets and
759 indications. *Nat Rev Drug Discov* 24, 458-479. 10.1038/s41573-025-01139-y.
- 760 62. Hauser, A.S., Attwood, M.M., Rask-Andersen, M., Schiöth, H.B., and Gloriam,
761 D.E. (2017). Trends in GPCR drug discovery: new agents, targets and
762 indications. *Nat Rev Drug Discov* 16, 829-842. 10.1038/nrd.2017.178.
- 763 63. Liskiewicz, D., Novikoff, A., Khalil, A., Akindehin, S., Campbell, J.E., Candela,
764 P., Castelino, R.L., Coupland, C., Culot, M., Dodson, W.S., et al. (2026).
765 GLP-1R–GIPR–PPAR α/δ quintuple agonism corrects obesity and diabetes
766 in mice. *Nature* 653, 776-785. 10.1038/s41586-026-10427-5.
- 767 64. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler,
768 H., Lewis, M., Yih, W.-t., and Rocktäschel, T. (2020). Retrieval-augmented
769 generation for knowledge-intensive nlp tasks. *Advances in neural information*
770 *processing systems* 33, 9459-9474.
- 771 65. Li, M., Kilicoglu, H., Xu, H., and Zhang, R. (2025). Biomedrag: A retrieval
772 augmented large language model for biomedicine. *Journal of Biomedical*

773 Informatics 162, 104769.

774

775

776 **Data availability**

777 The Obesity Knowledge Portal is freely accessible at <https://obesityknowledge.org>.
778 GWAS Catalog summary statistics used in this study were downloaded from the
779 NHGRI-EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas/downloads>). UK Biobank
780 summary statistics for BMI (Field 21001) and BMI by bioelectrical impedance (Field
781 23104) were obtained from the Neale Lab Round 2 GWAS results
782 (<https://www.nealelab.is/uk-biobank>). The RAG literature corpus was constructed from
783 publications indexed in PubMed Central (<https://www.ncbi.nlm.nih.gov/pmc/>).

784

785 **Acknowledgements**

786 We thank all members of the MultiOmics Lab for their contributions and discussions.
787 We thank the Neale Lab for making UK Biobank GWAS summary results publicly
788 available and the NHGRI-EBI GWAS Catalog team for their ongoing curation efforts.
789 This study was funded by the National Key Research and Development Program of
790 China (2024YFA1802800 and 2024YFA1802803), National Natural Science
791 Foundation of China under Grants (32371195 and W2541011), the Foundation of the
792 Shanghai Municipal Education Commission (24KXZNB10), and a new PI Start-up
793 grant of Fudan University to G.W. Computations were performed using the CFFF
794 platform of Fudan University.

795

796 **Competing Interests**

797 The authors declare no competing interests.

798

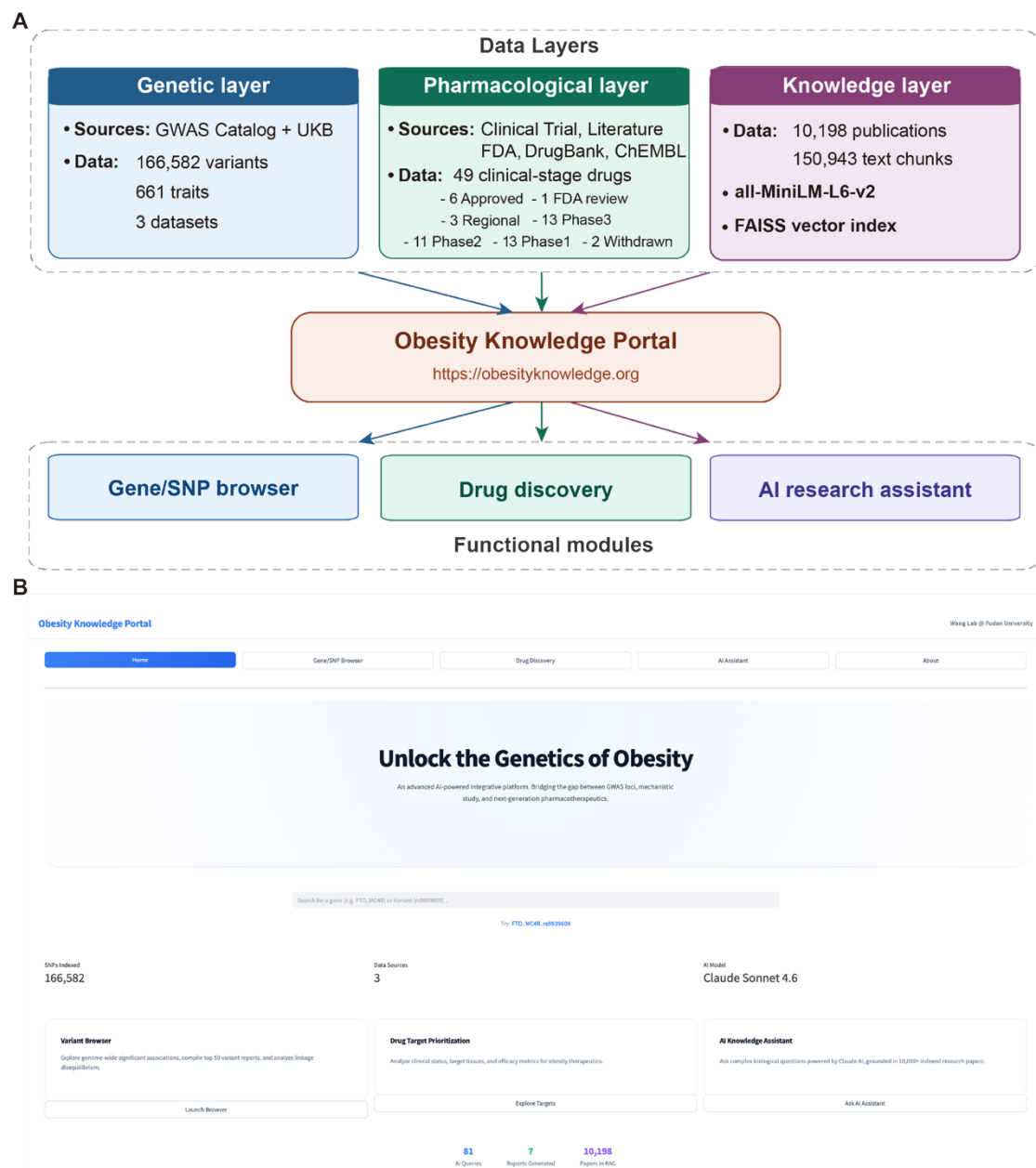
799 **Author Contributions**

800 Conceptualisation: G.W. & P.L.; Methodology: G.W. & Y.L. & W.X.; Software: G.W.
801 & Y.L.; Formal analysis: G.W. & Y.L.; Investigation: G.W. & Y.L. & B.L. & J.B.;

|

802 Data curation: G.W. & S.A.; Visualisation: G.W. & Y.L.; Writing, original draft: G.W.
803 & Y.L. & J.R.S.; Writing, review & editing: all authors; Supervision: G.W. & T.Z. &
804 J.R.S.; Funding acquisition: G.W. & J.R.S. All authors have read and approved the final
805 manuscript.
806

Figure 1



807

808 **Fig. 1 Overview of the Obesity Knowledge Portal.**

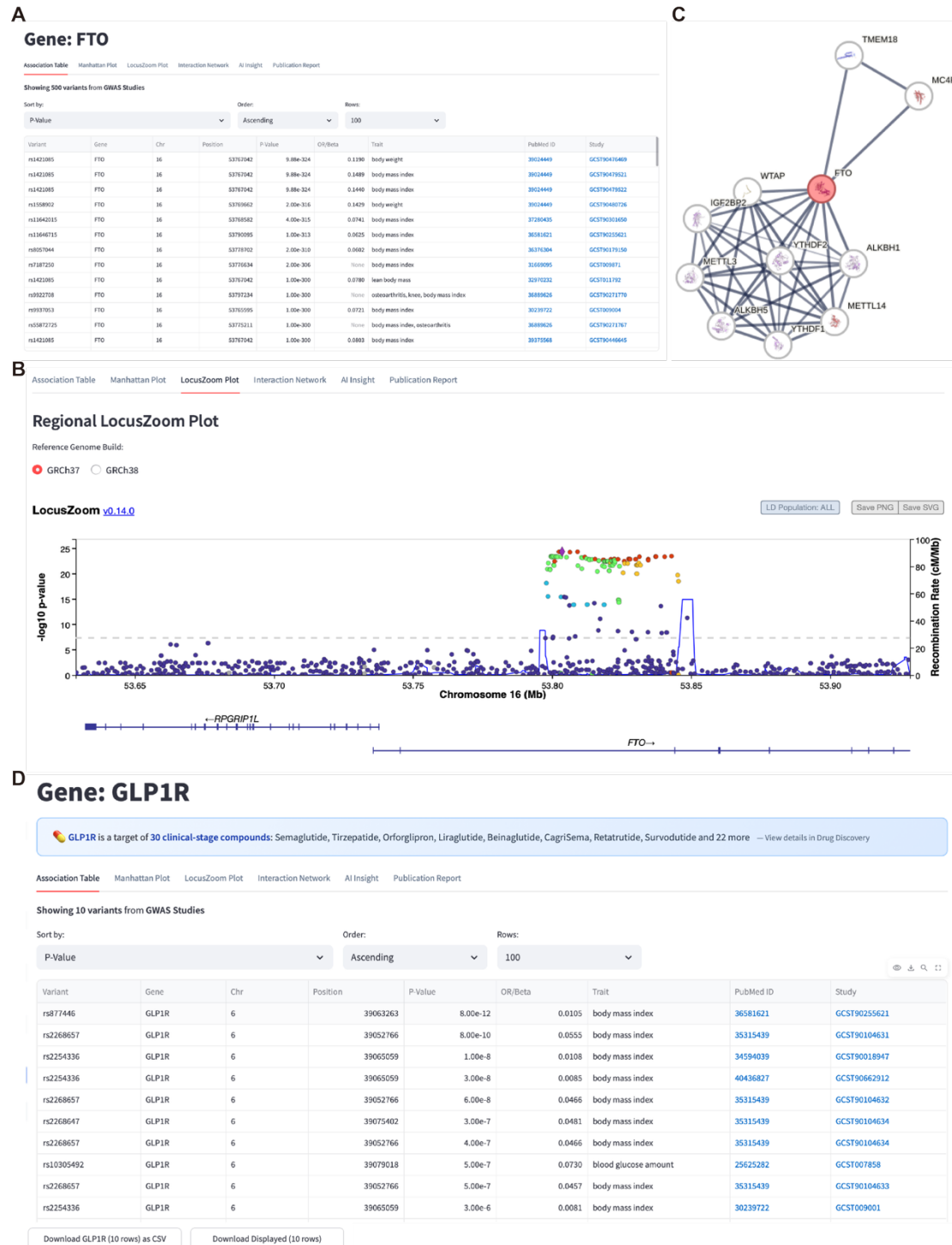
809 (A) Structure of the Obesity Knowledge Portal (OKP). The platform integrates three
810 complementary data layers: a genetic layer built on GWAS Catalog data and UK
811 Biobank data; a pharmacological layer comprising a manually curated database of 49
812 clinical-stage obesity therapeutics targeting 11 unique molecular targets; and a
813 knowledge layer built from 10,198 domain-specific publications segmented into
814 150,943 text chunks and indexed using FAISS for retrieval-augmented generation
815 (RAG). These data layers feed into three functional modules: an interactive

816 Gene/Variant Browser, a Drug Discovery module, and an AI Research Assistant
817 delivering literature-grounded, source-cited responses to genetic and pharmacological
818 queries.

819 (B) Home page of the OKP web interface (<https://obesityknowledge.org>), showing the
820 three functional modules.

821

Figure2



822

823 **Fig. 2 Gene/Variant Browser illustrated with two examples: *FTO* (A–C) and**
 824 ***GLP1R* (D).**

825 (A) Association table for *FTO* under the default GWAS dataset, returning 500 unique
 826 genome-wide significant variants on chromosome 16 (positions 53,750,466–
 827 54,117,066) across multiple obesity-related traits. Results are downloadable as CSV.

828 (B) *LocusZoom* regional association plot for the *FTO* locus (GRCh37), displaying
829 $-\log_{10}(P)$ values with linkage disequilibrium colouring derived from reference
830 populations and recombination rate overlay. Gene tracks for *FTO* and *RPGRIP1L* are
831 shown below.

832 (C) STRING protein–protein interaction network for the *FTO* protein, revealing
833 functional interaction partners including m⁶A RNA methylation complex members
834 (METTL3, METTL14, ALKBH5, YTHDF1/2) and obesity-associated proteins (MC4R,
835 TMEM18).

836 (D) Cross-linking of genetic and pharmacological evidence for *GLP1R*. The gene page
837 displays 10 genome-wide significant variants alongside a banner listing 30 clinical-
838 stage compounds targeting this receptor.

839

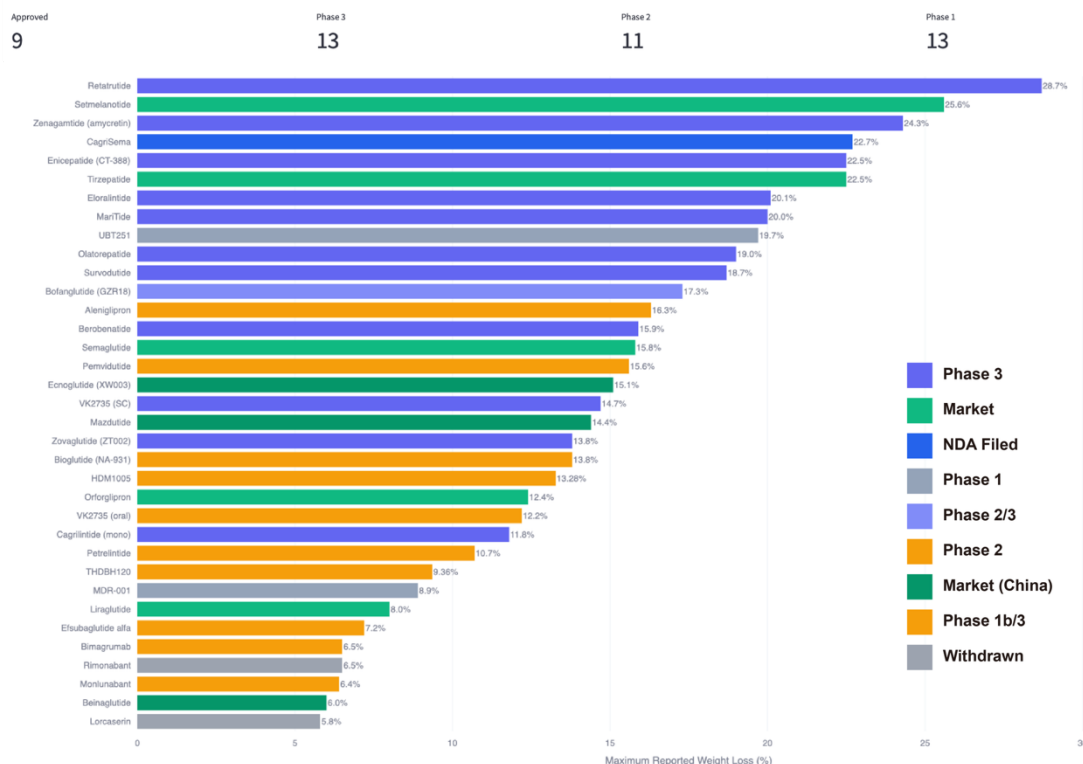
Figure3

A

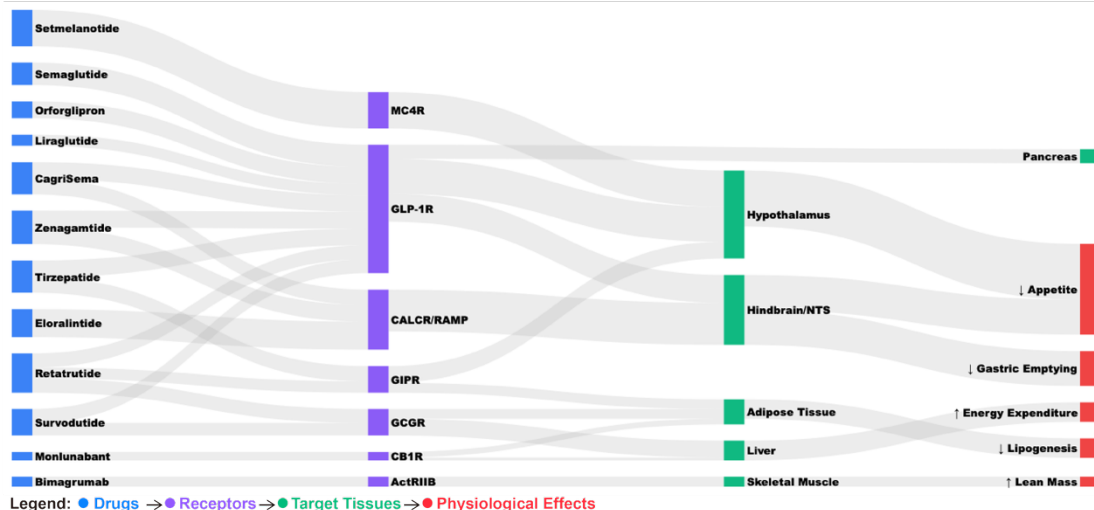
Obesity Therapeutics & Target Discovery

Pipeline Overview Drug Profiles Mechanisms & AI Novel Target Engine GWAS GPCRs Databases & References

Clinical Development Pipeline



B



840

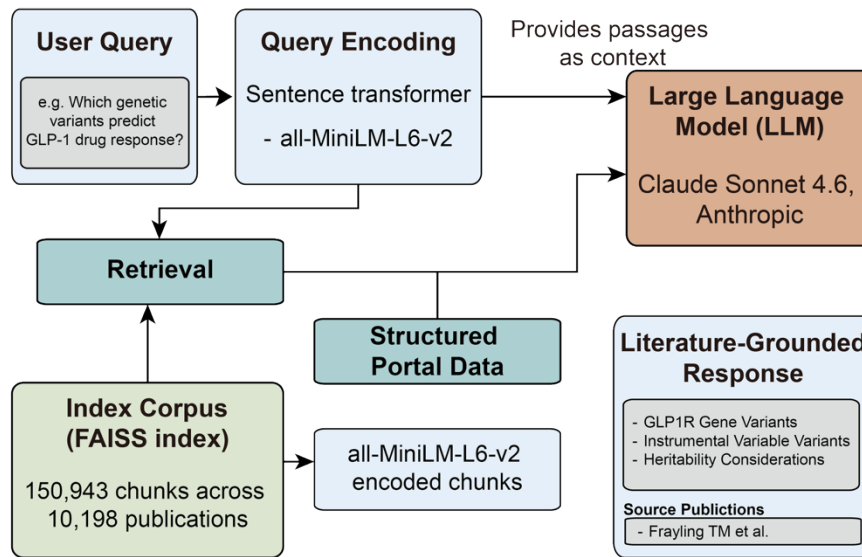
841 Fig. 3 Drug Discovery module.

842 (A) Pipeline Overview ranking all 49 clinical-stage obesity compounds by maximum
843 reported weight-loss efficacy (%), colour-coded by development stage.

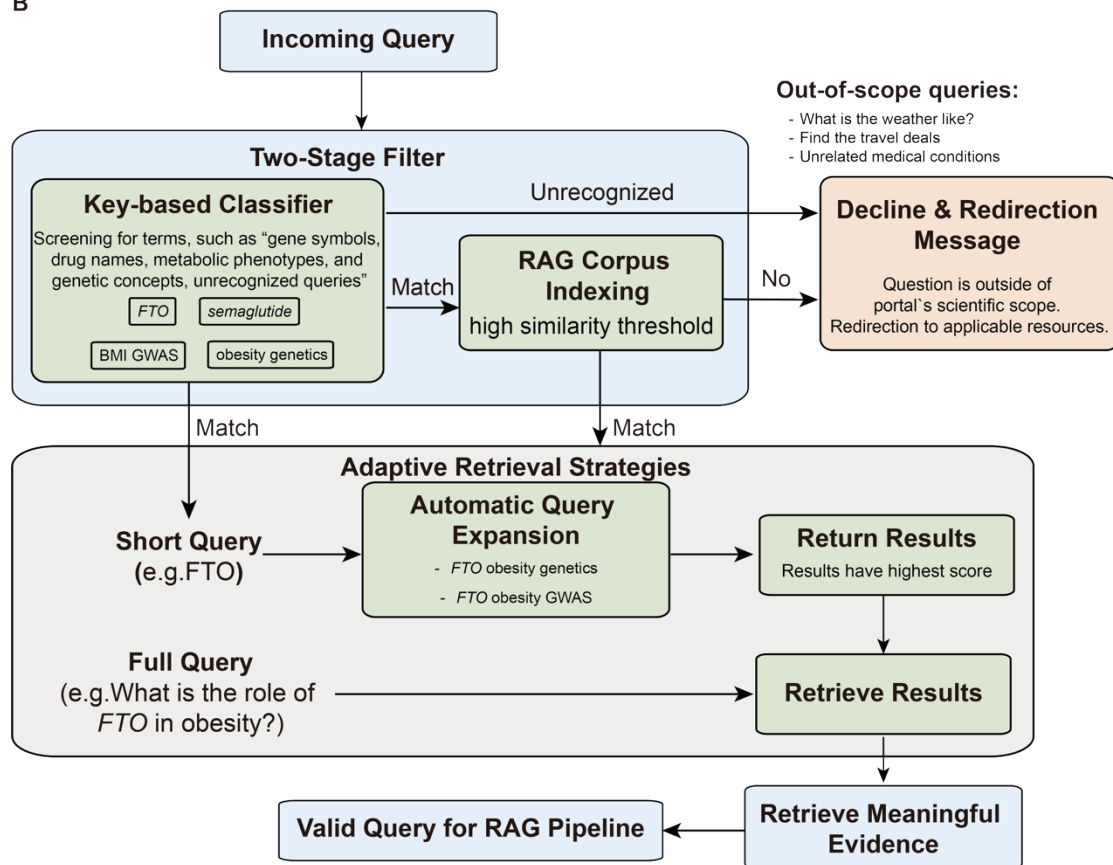
844 (B) Mechanisms & AI tab showing an interactive Sankey diagram mapping the flow

845 from drug compounds (left) through target receptors (*GLP1R*, *GIPR*, *CALCR/RAMP*,
846 *GCGR*, *CB1R*, *ActRIIB*, *MC4R*) to target tissues (hypothalamus, hindbrain/NTS,
847 pancreas, adipose tissue, liver, skeletal muscle) and physiological effects. Mechanistic
848 descriptions of four therapeutic paradigms including central appetite suppression,
849 energy expenditure, body composition preservation, and peripheral metabolic
850 modulation are provided below the diagram.

Figure4
A



B



851

852 **Fig. 4 AI-powered research assistant.**

853 (A) Architecture of the Retrieval-Augmented Generation (RAG) system. User queries

854 are encoded using the all-MiniLM-L6-v2 sentence transformer and matched against a

855 FAISS index of 150,943 text chunks from 10,198 publications. Retrieved passages,

856 together with structured portal data (GWAS associations and drug database), are
857 provided as context to a large language model (Claude Sonnet 4.6, Anthropic), which
858 generates a literature-grounded response accompanied by cited source publications
859 with links to PubMed Central.

860 (B) Query filtering and adaptive retrieval. A two-stage filter (keyword classifier and
861 RAG corpus similarity check) identifies obesity-relevant queries and declines out-of-
862 scope requests with a redirection message. Short or ambiguous inputs undergo
863 automatic query expansion to improve retrieval quality.

864

Fig S1

A

B

Gene: FTO

Association Table Manhattan Plot LocusZoom Plot Interaction Network **AI Insight** Publication Report

AI Insight

FTO – 500 variants in current dataset | Best P-value: $9.88e-324$ | Traits: smoking behavior, body mass index, body mass index, subcutaneous adipose tissue quality, waist-hip ratio, waist circumference ...

Generate RAG-Enhanced Summary

The FTO gene stands as one of the most robustly implicated loci in the genetics of human obesity, supported by an extraordinary breadth of genome-wide association evidence. Portal data for this gene reveals a dataset of 500 variants, of which 454 reach genome-wide significance at $P < 5e-8$, with a best association p-value of $9.88e-324$ – among the most statistically compelling signals ever observed in complex trait genetics. The associated traits span a remarkably wide phenotypic spectrum, encompassing body mass index, body fat percentage, waist circumference, hip circumference, waist-hip ratio, visceral and subcutaneous adipose tissue quality, leptin measurement, metabolic syndrome, and obesity disorder itself, as well as longitudinal BMI measurements and obese and overweight BMI status categories. Beyond adiposity per se, FTO variants have been associated with cardiometabolic comorbidities including type 2 diabetes, coronary artery disease, heart failure, stroke, atrial fibrillation, hypertension metrics such as systolic and diastolic blood pressure, and lipid traits. Associations have also been reported with polycystic ovary syndrome, breast cancer in interaction with adiposity-related phenotypes, osteoarthritis of the hip and knee, bone mineral density, and even behavioral phenotypes such as smoking behavior, energy intake, physical activity, and diet measurement. This extraordinary trait pleiotropy underscores FTO as a central node in the genetic architecture of metabolic and cardiometabolic disease.

The discovery of FTO's association with BMI and predisposition to both childhood and adult obesity represented a landmark in complex disease genetics. Subsequent independent studies confirmed that variation in FTO contributes to a broad continuum of adiposity, from overweight to severe and morbid obesity, and that the effect is detectable across diverse ancestral populations. Meta-analytic evidence has further confirmed that specific FTO polymorphisms, particularly those in the intronic region of the gene, confer consistent and reproducible risk for obesity susceptibility in both children and adults. Importantly, research has demonstrated that the impact of FTO genetic variants on obesity and its metabolic consequences is not fixed but can be modulated by environmental factors, particularly daily macronutrient intake, highlighting a meaningful gene-environment interaction that may be relevant for precision nutrition approaches.

At the molecular level, FTO encodes a 2-oxoglutarate-dependent nucleic acid demethylase, establishing it as an RNA demethylase that catalyzes the removal of N6-methyladenosine (m6A) modifications from messenger RNA. This enzymatic function places FTO at the intersection of epitranscriptomics and metabolic regulation, as m6A modifications are now recognized as pervasive regulators of mRNA stability, translation efficiency, and splicing. Experimental evidence has identified small molecule inhibitors of FTO's demethylase activity – for example, entacapone has been characterized as a chemical inhibitor of FTO that mediates metabolic regulation through the transcription factor FOXO1, suggesting that pharmacological targeting of FTO's catalytic activity is chemically feasible. Furthermore, research has implicated FTO variants in the relationship between obesity and depression through environmental interactions that may influence overlapping neurobiological pathways, adding a neuropsychiatric dimension to its biology.

From a therapeutic standpoint, the convergence of genetic, biochemical, and pharmacological evidence positions FTO as a compelling drug target in metabolic disease. The gene's association with visceral adipose tissue quantity, abdominal adipose tissue, body fat distribution, and gluteofemoral adiposity suggests that modulation of FTO activity could influence not just total body weight but also the depot-specific distribution of fat, which is a key determinant of cardiometabolic risk independent of BMI. The breadth of FTO's association landscape – extending to insulin-related traits, lipid parameters including HDL cholesterol, LDL triglycerides, and total cholesterol, as well as sex hormone-binding globulin – further suggests that therapeutic interventions targeting this pathway could have systemic metabolic benefits beyond weight reduction alone. As the field of anti-obesity pharmacotherapy continues to advance, FTO's well-characterized genetic signal, its biochemically tractable enzymatic function, and its pleiotropic reach across adiposity and cardiometabolic traits collectively make it a high-priority candidate for translational investigation.

References

1. Feysing M et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* (2002)358:949–954 (2002). PMID: 1244486
2. Linn S, Yan C-H. The genetics of obesity from discovery to biology. *Nat Rev Genet* (2023)23:133–153 (2023). PMID: 3698890
3. Lohman M et al. Genetic studies of body mass index yield new insights for biology. *Nature* (2015)518:245–250 (2015). PMID: 26174432
4. Spadoni R et al. Association analyses of 247,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* (2012)44:977–983 (2012). PMID: 22900081
5. Yang J et al. Meta-analysis of genome-wide association studies for height and body mass index in 790,000 individuals. *Nat Genet* (2010)42:561–569 (2010). PMID: 20124942
6. Lohman M et al. Genetics of Common Obesity in Children and Adolescents. *Annals of the New York Academy of Sciences* (2015). PMID: 4188226
7. Du M et al. Cross-trait genomic modeling reveals the pleiotropic architecture and systemic impact of MBO. *Science advances* (2020). PMID: 4038886
8. Rhee SH et al. The Role of Adipocytes and Gene Polymorphisms in the Development of Obesity-Induced Depression. *Current obesity reports* (2023). PMID: 40720866
9. Fang J et al. Obesity and emerging intervention strategies: Mechanistic insights and novel drug targets. *Chinese medical journal* (2020). PMID: 41286252
10. Kang H et al. Obesity pathophysiology and therapeutic interventions. *Molecular biomedicine* (2023). PMID: 40278965
11. Green D et al. Fine-Scale Haplotype Mapping Reveals an Association of the FTO Gene with Osteoporosis and Fracture Risk in Postmenopausal Women. *Genes* (2024). PMID: 39326742
12. Imegiayana A et al. Genetic Insights into Severe Obesity: A Case Study of MC4R Variant Identification and Clinical Implications. *Genes* (2023). PMID: 40483251
13. Raqeeb H et al. Association of anthropometric indices with rs936989 FTO gene polymorphism among overweight/obese women with breast cancer: A case-control study. *Frontiers in nutrition* (2023). PMID: 40381633
14. Hajj-Ismail Ghazali M et al. Interaction of common variants of FTO gene and Dietary Inflammatory Index on obesity measures: Tehran Lipid and Glucose Study. *BMJ nutrition, prevention & health* (2023). PMID: 38118282
15. Neavil E et al. Genetic: A Starting Point for the Prevention and the Treatment of Obesity. *Nutrients* (2023). PMID: 37173686
16. Hossaini Edghani F et al. Dietary approach to sleep hypertension and healthy eating index 2015, modify the association between FTO polymorphisms and obesity phenotypes. *BMC research notes* (2023). PMID: 3787388
17. Song T et al. FTO rs936989 and rs1761449 polymorphisms contribute to metabolic syndrome risk by increasing triglyceride and glucose levels. *Frontiers in genetics* (2023). PMID: 4144066
18. Maldonado-Arauz JE et al. The transition from preclinical to clinical obesity: the importance of a borderline stage. *Clinical science (London, England: 1875-)* (2020). PMID: 4146666
19. Perrot S et al. Association of FTO variants rs936989 and rs1420205 with elevated sugar and fat consumption in adult obesity. *Scientific reports* (2024). PMID: 3943444
20. Khamarovsky Nuhman L et al. The Common FTO rs936989 Polymorphism Interacts with Sleeping and Eating Habits to Affect Fasting Glucose in Type 2 Diabetes. *Nutrients* (2024). PMID: 41463220
21. Han Y et al. In Vivo Genetic Expression of FTO and MC4R in Stomach Gastric Patients. *Diagnostic Utility Without Predictive Value for Weight Loss. Obesity surgery* (2024). PMID: 41363115
22. Veendharna A et al. The Genetic Basis of Childhood Obesity: A Systematic Review. *Nutrients* (2023). PMID: 3886146
23. Liu Y et al. A Complex Network of Obesity Risk Genes Revealed by Systematic Bioinformatics and Single-Cell Transcriptomic Analyses. *Journal of obesity* (2025). PMID: 40201885

865

866 **Supplemental Fig. 1 Gene/Variant Browser extended views.**

867 (A) Overview of the browser interface. The Configuration panel allows dataset

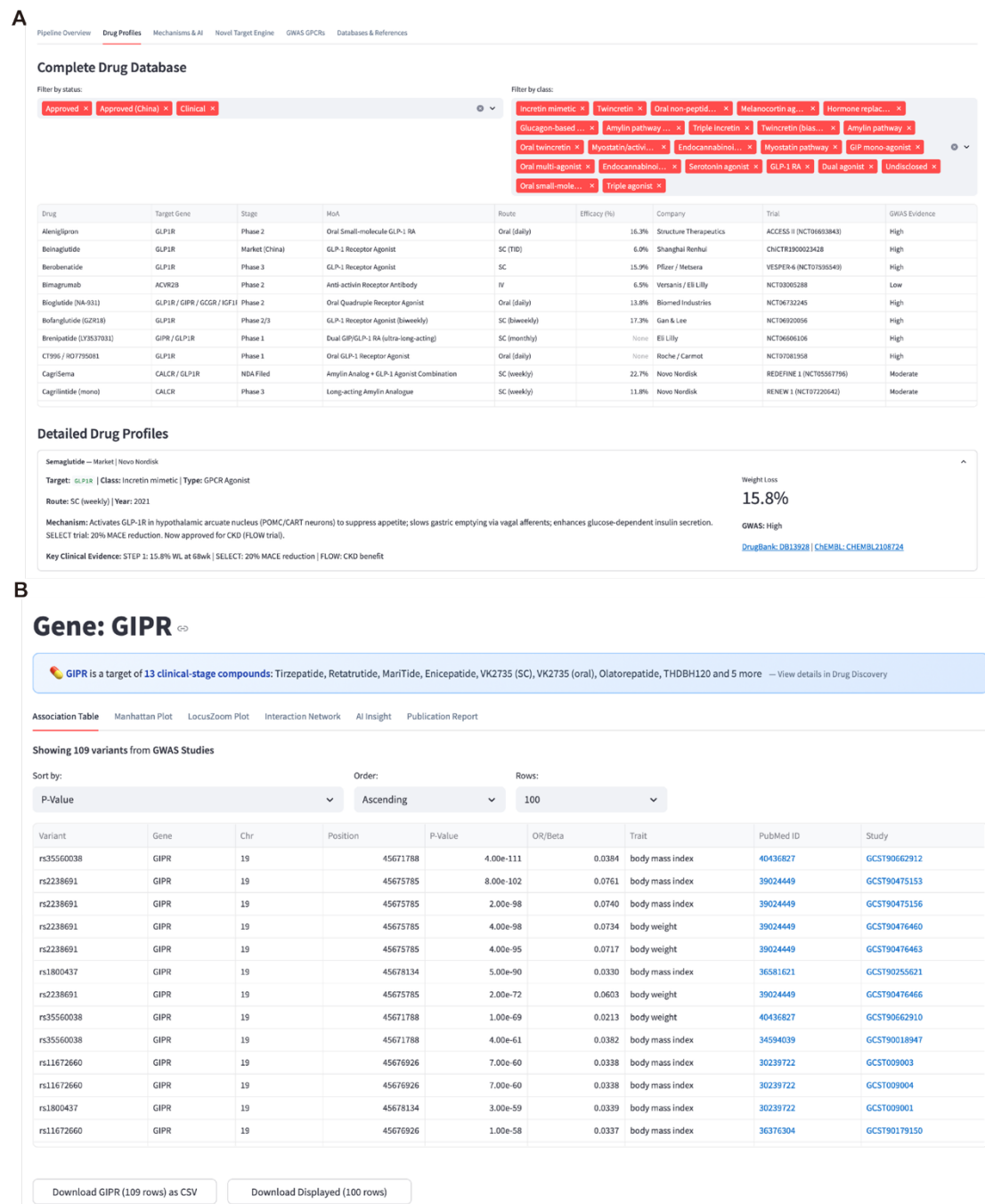
868 selection (GWAS, UK Biobank BMI [Field 21001], or UK Biobank BMI Impedance

869 [Field 23104]), and the Filter panel supports gene-, variant-, and trait-level filtering
870 across obesity-related phenotypes. Six analytical tabs are available: (1) Association
871 Table, (2) Manhattan Plot, (3) LocusZoom Plot, (4) Interaction Network, (5) AI Insight,
872 and (6) Publication Report. The association table is sortable and filterable, with direct
873 links to PubMed records and GWAS Catalog study accessions.

874 (B) AI Insight tab for the *FTO*, generating a literature-grounded summary covering
875 biological function, GWAS associations, molecular mechanisms, and therapeutic
876 potential, with cited source publications and PubMed links shown in the References
877 panel.

878

Fig S2



879

880 **Supplemental Fig. 2 Drug Discovery module and genetic evidence for the GIPR**
 881 **incretin target.**

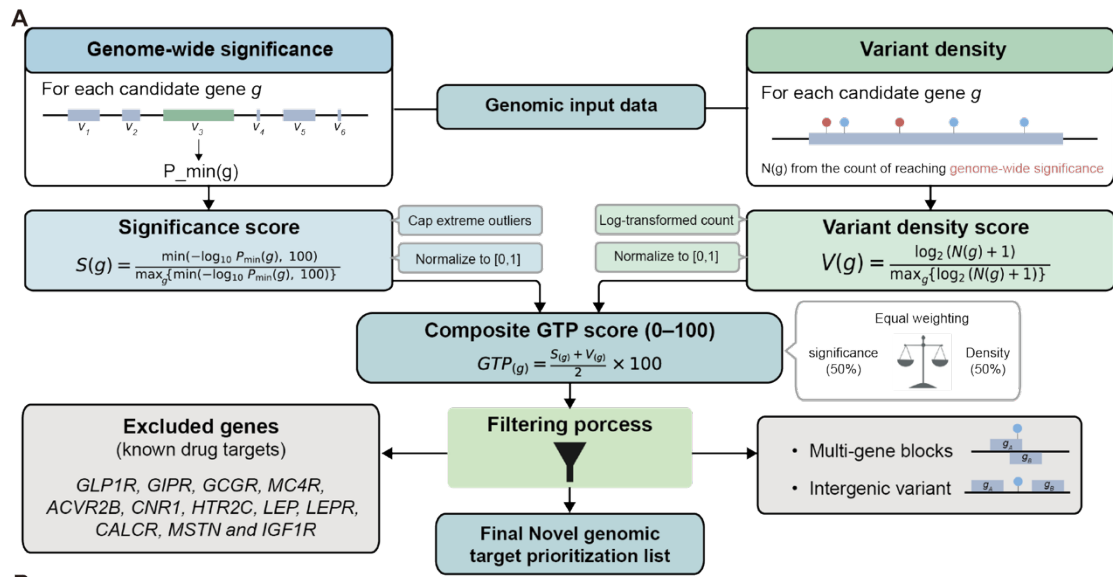
882 (A) Drug Profiles tab, showing the Complete Drug Database—a filterable table of the
 883 curated obesity therapeutics with status- and class-based filters, together with the
 884 Detailed Drug Profiles panel, which expands to summarise the key clinical evidence
 885 for each compound (Semaglutide shown as an example), with external links to

886 DrugBank and ChEMBL.

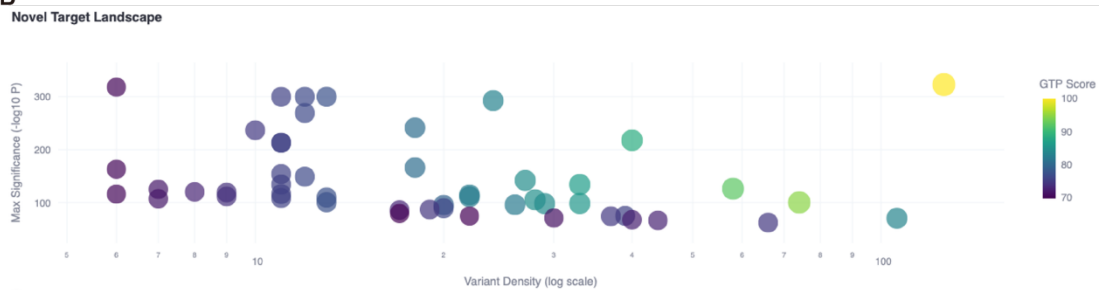
887 (B) Gene/Variant Browser view of GIPR under the GWAS dataset, returning 109
888 variants. A cross-linking banner indicates that GIPR is targeted by 13 clinical-stage
889 compounds. The association table is sortable by P value, chromosomal position, or
890 effect size, with direct links to PubMed records and GWAS Catalog study accessions,
891 and is exportable as CSV.

892

Fig S3



B



C

Select a high-scoring novel target for AI viability report:

FTO Generate Viability Report

AI Viability Report:
FTO (Fat Mass and Obesity-Associated protein) has emerged as one of the most compelling and extensively studied genetic loci in human obesity research, making it a highly relevant candidate for therapeutic targeting. Understanding its biological role, druggability, expression landscape, and the current state of pharmacological intervention is essential for evaluating its potential as a novel anti-obesity target.

Biological Function and Pathway
FTO encodes a 2-oxoglutarate-dependent nucleic acid demethylase, placing it within the RNA modification machinery rather than the classical appetite-signaling cascades. Specifically, the protein catalyzes the oxidative demethylation of N6-methyladenosine (m6A) residues in RNA, a post-transcriptional modification with broad regulatory consequences for mRNA stability, splicing, and translation. Through this epitranscriptomic mechanism, FTO influences pathways governing energy intake and adipocyte differentiation. Portal data from the Obesity Knowledge Portal identifies FTO as a gene with a minimum association p-value of $1e-323$ and 126 significant variants at genome-wide significance thresholds ($P < 5 \times 10^{-8}$), underscoring the extraordinary strength and consistency of its genetic signal in obesity GWAS. Beyond its primary role in energy homeostasis, research has linked FTO to a broader set of metabolic phenotypes including type 2 diabetes (T2D), non-alcoholic fatty liver disease (NAFLD), and hypertension. Genome-wide analyses have further identified pleiotropic effects of the FTO locus on both obesity and osteoporosis, suggesting that its regulatory influence extends across multiple metabolic and tissue-maintenance programs. Intriguingly, variation in FTO has also been implicated in pathways connecting obesity with depression, potentially through gene-environment interactions that modulate neurobiological circuits governing mood and energy balance simultaneously.

GPCR Status and Druggability
Portal annotations confirm that FTO is not a G-protein coupled receptor (GPCR). Instead, it belongs to the AlkB family of non-heme iron- and 2-oxoglutarate-dependent dioxygenases. This enzymatic classification is actually favorable from a drug discovery perspective, as the active site harbors a well-defined catalytic pocket that coordinates iron and 2-oxoglutarate, providing a structurally tractable binding site for small-molecule inhibitors. Enzyme active sites of this class are in principle amenable to competitive or allosteric inhibition with drug-like molecules, and the availability of structural data on the FTO protein has facilitated rational drug design efforts. While it is not a GPCR and therefore cannot be targeted via the well-trodden receptor agonist/antagonist modalities used for targets such as GLP-1R or MC4R, its enzymatic nature offers a distinct and pharmacologically exploitable mechanism of action.

Tissue Expression
FTO exhibits broad tissue expression, and its functional relevance varies considerably across compartments. In adipose tissue, FTO overexpression has been consistently observed in obese individuals prior to therapeutic intervention, aligning with its established role in regulating adipocyte differentiation and lipid accumulation. Clinical studies have demonstrated that obese participants overexpress FTO at baseline, and that therapeutic interventions capable of downregulating FTO expression correlate with improvements in metabolic parameters, suggesting that adipose tissue represents a particularly important site of action. In the brain, FTO is expressed in hypothalamic nuclei involved in appetite regulation, and central FTO activity is thought to contribute to the neurological control of food intake and energy expenditure. This dual expression in brain and adipose tissue positions FTO at the intersection of central appetite signaling and peripheral metabolic regulation. Expression in liver and skeletal muscle has been noted in the context of broader metabolic disease associations such as T2D and NAFLD, though these tissues are considered secondary sites relative to the brain-adipose axis for the core obesity phenotype.

Existing Compounds and Tool Molecules
Pharmacological targeting of FTO has progressed meaningfully at the preclinical level. Research has identified entacapone — a catechol-O-methyltransferase inhibitor previously approved for Parkinson's disease — as a chemical inhibitor of FTO that mediates metabolic regulation through the transcription factor FOXO1. This repurposing finding is significant because it demonstrates that a clinically validated, orally bioavailable small molecule can engage the FTO active site and produce downstream metabolic effects in vivo. Beyond entacapone, a growing suite of tool compounds

893

894 **Supplemental Fig. 3 Genomic target prioritisation engine.**

895 (A) Overview of the Novel Target Discovery Engine, introducing the Genomic Target

896 Prioritisation (GTP) Score.

897 (B) Novel Target Landscape scatter plot displaying all scored candidate genes by
898 variant density (x-axis, log scale) against maximum statistical significance ($-\log_{10}P$, y-
899 axis), with bubble colour encoding GTP score.

900 (C) AI Viability Report generated for *FTO*, providing a literature-grounded assessment
901 of biological function, druggability, tissue expression, existing tool molecules, and drug
902 development challenges. Retrieved source references with relevance scores are
903 displayed below the report for qualification.

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

|

Fig S4

A
AI Assistant

The screenshot shows an AI Assistant interface. At the top right, a blue button contains the query: "How does FTO influence obesity through downstream regulatory genes?". Below this, a text box contains the AI's response. The response is structured into several paragraphs: a general overview of the FTO gene's role in obesity genetics; a detailed description of its mechanism as an m6A RNA demethylase; a discussion of its physiological effects on appetite and energy balance; evidence from animal models; and a mention of environmental interactions. Below the response is a text input field with the placeholder "e.g. What are the strongest GWAS signals for BMI?" and a "Send" button. At the bottom, a "References" section lists 27 numbered citations with their titles and PMIDs.

How does FTO influence obesity through downstream regulatory genes?

The FTO gene represents one of the most extensively studied loci in human obesity genetics, and considerable research has shed light on how it exerts its effects through downstream regulatory mechanisms rather than solely through direct protein function. Genome-wide association studies have robustly identified common variants within the FTO locus — particularly rs9939609 and related single nucleotide polymorphisms — as strongly associated with body mass index and predisposition to both childhood and adult obesity. Importantly, the mechanistic basis of this association has become clearer over time: the obesity-risk variants in FTO appear to act largely through long-range chromatin interactions that regulate the expression of downstream genes, most notably IRX3 and IRX5, which are transcription factors involved in adipocyte thermogenesis and energy homeostasis. Disruption of this regulatory axis has been shown to shift the balance of adipocyte differentiation toward energy-storing white fat and away from thermogenic beige or brown adipocyte programs, thereby promoting fat accumulation.

At the behavioral and physiological level, studies in individuals carrying FTO risk alleles indicate that the increase in BMI is primarily driven by elevated energy intake and reduced feelings of food satiety, rather than by a marked reduction in energy expenditure per se. This points to regulatory influences on appetite-controlling pathways, potentially involving hypothalamic circuits that govern hunger signaling.

Evidence from animal models further substantiates the downstream regulatory role of FTO. Loss of FTO function in mice resulted in reduced body weight and lower fat mass, while overexpression produced the opposite phenotype, confirming that FTO dosage has meaningful consequences for adiposity through the regulatory networks it controls.

Beyond energy balance, research has also explored how FTO genetic variants interact with environmental factors — particularly macronutrient intake — to modulate metabolic outcomes, suggesting that the downstream regulatory effects of FTO are not static but are shaped by gene–environment interactions. Additionally, FTO variants have been implicated in shared pathways connecting obesity with conditions such as depression, further illustrating the breadth of downstream regulatory influence this locus exerts across multiple physiological systems.

e.g. What are the strongest GWAS signals for BMI?

References

1. Frayling TM et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* (316):889–894 (2007). [PMID: 17434869]
2. Loos RJF, Yeo GSH. The genetics of obesity: from discovery to biology. *Nat Rev Genet* (23):129–133 (2022). [PMID: 3455834]
3. Locke AE et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* (518):75–82 (2015). [PMID: 25673433]
4. Speliotes EK et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* (42):11–15 (2010). [PMID: 20939430]
5. Yengo L et al. Meta-analysis of genome-wide association studies for height and body mass index. *Hum Mol Genet* (27):205–216 (2018). [PMID: 30124472]
6. Drucker DJ. GLP-1-based therapies for diabetes, obesity and beyond. *Nat Rev Drug Discov* (24):631–650 (2025). [PMID: 40281394]
7. Wilding JPH et al. Once-weekly semaglutide in adults with overweight or obesity (STEP 1). *NEJM* (384):1196–1302 (2021). [PMID: 33547185]
8. Jastreboff AM et al. Tirzepatide once weekly for the treatment of obesity (SURMOUNT-1). *NEJM* (387):205–216 (2022). [PMID: 3554624]
9. Kong Y et al. Obesity: pathophysiology and therapeutic interventions. *Molecular biomedicine* (2025). [PMID: 40278940]
10. Naqeeb H et al. Association of anthropometric indices with rs9939609 FTO gene polymorphism among overweight/obese women with breast cancer: a case-control study. *Frontiers in nutrition* (2025). [PMID: 40636185]
11. León-Reyes G et al. Genetics of Common Obesity in Children and Adolescents. *Annals of the New York Academy of Sciences* (2025). [PMID: 41082230]
12. Ercan SN et al. The Role of Adipokines and Gene Polymorphisms in the Development of Obesity-Induced Depression. *Current obesity reports* (2025). [PMID: 40720046]
13. Benak O et al. FTO in health and disease. *Frontiers in cell and developmental biology* (2024). [PMID: 39744011]
14. Du M et al. Cross-trait genomic modeling reveals the polygenic architecture and systemic impact of MASLD. *Science advances* (2026). [PMID: 41484896]
15. Nowvili G et al. Genetics: A Starting Point for the Prevention and the Treatment of Obesity. *Nutrients* (2023). [PMID: 37375486]
16. Maldonado-Ariza JE et al. The transition from preclinical to clinical obesity: the importance of a borderline stage. *Clinical science (London, England)* : 1979 (2020). [PMID: 41496666]
17. Hajj-Hosseini-Gazestani N et al. Interaction of common variants of FTO gene and Dietary Inflammatory Index on obesity measures: Tehran Lipid and Glucose Study. *BMJ nutrition, prevention & health* (2022). [PMID: 3861829]
18. Hany M et al. In Vivo Gastric Expression of FTO and MC4R in Sleeve Gastrectomy Patients: Diagnostic Utility Without Predictive Value for Weight Loss. *Obesity surgery* (2025). [PMID: 41361717]
19. Górczyńska-Kosińska S et al. Exploring the Interplay of Genetics and Nutrition in the Rising Epidemic of Obesity and Metabolic Diseases. *Nutrients* (2024). [PMID: 39458554]
20. Bosconi S et al. Obesity: a disease of the ponderostat and the regulation of energy balance. *Eating and weight disorders : EWD* (2025). [PMID: 41091421]
21. Nikolaeva AF et al. When Genes Wear Masks: Epigenomic Modulation in the Development and Progression of Obesity. *International journal of molecular sciences* (2025). [PMID: 40809388]
22. Greer D et al. Fine-Scale Haplotype Mapping Reveals an Association of the FTO Gene with Osteoporosis and Fracture Risk in Postmenopausal Women. *Genes* (2024). [PMID: 39336140]
23. Imangaliyeva A et al. Genetic Insights into Severe Obesity: A Case Study of MC4R Variant Identification and Clinical Implications. *Genes* (2025). [PMID: 40438329]
24. Feng X et al. Obesity and emerging intervention strategies: Mechanistic insights and novel drug targets. *Chinese medical journal* (2026). [PMID: 41394202]
25. Kazanroosky Nahshan L et al. The Common FTO rs9939609 Polymorphism Interacts with Sleeping and Eating Windows to Affect Predisposition to Type 2 Diabetes. *Nutrients* (2026). [PMID: 41683295]
26. Al-Serri A et al. A Cross-Sectional Study Investigating the Relationship Between the FTO Gene Polymorphism in Relation to Obesity Traits and Vitamin D Status in Adolescence. *Obesity science & practice* (2026). [PMID: 41523363]
27. Trang K et al. Genetics and epigenetics in the obesity phenotyping scenario. *Reviews in endocrine & metabolic disorders* (2023). [PMID: 37321403]

922

923 **Supplemental Fig. 4. Example output from the OKP AI assistant.**

924 (A) The AI Assistant panel responding to the user query "How does FTO influence
925 obesity through downstream regulatory genes?". The response describes FTO's
926 function as an m⁶A RNA demethylase regulating the stability and translation of
927 metabolic transcripts, its action on appetite-regulating genes in the hypothalamus, and
928 its long-range enhancer-mediated regulation of downstream genes such as IRX3 and
929 IRX5, which govern thermogenesis and adipocyte differentiation.

930

Table 1. Comprehensive obesity therapeutic pipeline integrated into the Obesity Knowledge Portal.

Drug	Target	Mechanism	Stage	Trial ID(s)	WL%	GWAS Ev.	Company	Route
FDA-Approved								
Semaglutide	GLP-1R	GLP-1 receptor agonist	Approved	STEP 1 (NCT03548935)	15.8	High	Novo Nordisk	SC QW
Tirzepatide	GIPR/GLP-1R	Dual GIP/GLP-1 RA	Approved	SURMOUNT-1 (NCT04184622)	22.5	High	Eli Lilly	SC QW
Orforglipron	GLP-1R	Oral non-peptide GLP-1 RA	Approved (Apr 2026)	ATTAIN-1 (NCT05051579)	12.4	High	Eli Lilly	PO QD
Liraglutide	GLP-1R	GLP-1 receptor agonist	Approved	SCALE (NCT01272219)	8.0	High	Novo Nordisk	SC QD
Setmelanotide	MC4R	MC4R agonist	Approved (monogenic)	NCT02896192	25.6	Very High	Rhythm Pharma	SC QD
Metreleptin	LEPR	Recombinant leptin	Approved (lipodystrophy)	NIH trials	N/A	High	Amryt Pharma	SC QD
Regionally Approved								
Mazdutide	GCGR/GLP-1R	Dual glucagon/GLP-1 RA	China NMPA (Jun 2025)	GLORY-1 (NCT05607680)	14.4	Moderate	Innovent Biologics	SC QW
Ecnoglutide (XW003)	GLP-1R	cAMP-biased GLP-1 RA	China NMPA (Mar 2026)	SLIMMER (NCT05813795)	15.1	High	Sciwind Biosciences	SC QW
Beinaglutide	GLP-1R	GLP-1 receptor agonist	China NMPA-approved (T2D, 2016); Phase 3 completed (obesity) ChiCTR1900023428	NCT03829891 ChiCTR1900023428	6.0	High	Shanghai Renhui	SC TID
Under FDA Review								
CagriSema	AMY-R/GLP-1R	Amylin + GLP-1 RA combo	NDA filed (Dec 2025)	REDEFINE 1 (NCT05567796)	22.7	Moderate	Novo Nordisk	SC QW
Phase 3								
Retatrutide	GIPR/GLP-1R/GCGR	Triple agonist	Phase 3	TRIUMPH-4 (NCT05931367)	28.7	Moderate	Eli Lilly	SC QW
Survodutide	GCGR/GLP-1R	Dual glucagon/GLP-1 RA	Phase 3	SYNCHRONIZE-1 (NCT06066528)	18.7	Moderate	BI / Zealand	SC QW
MariTide	GIPR(ant)/GLP-1R	GIP antagonist + GLP-1 ag.	Phase 3	MARITIME-1 (NCT06858839)	20.0	Moderate	Amgen	SC QM
Enicepatide (CT-388)	GIPR/GLP-1R	Biased dual GLP-1/GIP RA	Phase 3	Enith1 (NCT07351045)	22.5	High	Roche / Carmot	SC QW
VK2735 (SC)	GIPR/GLP-1R	Dual GLP-1/GIP RA	Phase 3	VANQUISH-1 (NCT06616870)	14.7 (13wk)	Moderate	Viking Therapeutics	SC QW
Zenagamtide (amycretin)	AMY-R/GLP-1R	Unimolecular GLP-1/amylin	Phase 3	AMAZE 1 (NCT07339423)	24.3	Moderate	Novo Nordisk	SC QW
Eloralintide	AMY-R	Selective amylin RA	Phase 3	NCT07321886	20.1	Moderate	Eli Lilly	SC QW
Cagrilintide (mono)	AMY-R	Long-acting amylin analogue	Phase 3	RENEW 1 (NCT07220642)	11.8	Moderate	Novo Nordisk	SC QW
Zovaglutide (ZT002)	GLP-1R	Long-acting GLP-1 RA (monthly)	Phase 3	HORIZON-1 (NCT07230119)	13.8	High	QL Biopharm	SC QM
Olatorepatide	GIPR/GLP-1R	Dual GLP-1/GIP RA	Phase 3 (CN done)	NCT07431086	19.0	Moderate	Hansoh / Regeneron	SC QW
Berobenatide	GLP-1R	GLP-1 receptor agonist	Phase 3	VESPER-6 (NCT07595549)	15.9	High	Pfizer / Metsera	SC QW/QM

Efsubaglutide alfa	<i>GLP-1R</i>	GLP-1 RA	Phase 2b/3	NCT06921486	7.2	High	Innogen	SC QW
Bofaglutide (GZR18)	<i>GLP-1R</i>	GLP-1 RA (biweekly)	Phase 2/3	NCT06737042 (phase 2 US); NCT06728124 (phase 3 China)	17.3	High	Gan & Lee	SC Q2W/QM
Phase 2								
Aleniglipron	<i>GLP-1R</i>	Oral small-molecule GLP-1 RA	Phase 2	ACCESS II (NCT06693843)	16.3	High	Structure Therapeutics	PO QD
VK2735 (oral)	<i>GIPR/GLP-1R</i>	Oral dual GLP-1/GIP RA	Phase 2	VENTURE-Oral (NCT06828055)	12.2 (13wk)	Moderate	Viking Therapeutics	PO QD
Bimagrumb	<i>ActRIIA/B</i>	Anti-activin receptor mAb	Phase 2	NCT03005288	6.5 (BW)	Low	Versanis / Lilly	IV
Monlunabant	<i>CB1R</i>	Peripheral CB1 inverse agonist	Phase 2	NCT05891834	6.4	High	Novo Nordisk	PO
Petrelintide	<i>AMY-R</i>	Long-acting amylin analogue	Phase 2	ZUPREME-1 (NCT06662539)	10.7	Moderate	Zealand / Roche	SC QW
Pemvidutide	<i>GCGR/GLP-1R</i>	Dual glucagon/GLP-1 RA	Phase 2	MOMENTUM (NCT05295875)	15.6	Moderate	Altimmune	SC QW
Taldefgrobep alfa	<i>Myostatin</i>	Anti-myostatin mAb	Phase 2b	NCT07281495	TBD	Low	Biohaven	SC
THDBH120	<i>GIPR/GLP-1R</i>	Dual GIP/GLP-1 receptor agonist (ultra-long-acting)	Phase 2	NCT07036601	9.36	High	Tonghua Dongbao	SC QW/Q2W
HDM1005	<i>GIPR/GLP-1R</i>	Dual GIP/GLP-1 receptor agonist	Phase 2	NCT07279194 (CN); NCT06886126 (US)	13.28	—	Huadong Medicine	SC QW
Macupatide	<i>GIPR</i>	GIP receptor agonist	Phase 2	NCT07589608	—	High	Eli Lilly	SC
Bioglutide (NA-931)	<i>IGF-1R/GLP-1R/GIPR/GCGR</i>	First-in-class oral quadruple receptor agonist	Phase 2	NCT06732245	13.8	High (GLP1R, GIPR, GCGR)	Biomed Industries	PO QD
Phase 1								
Brenipatide (LY3537031)	<i>GIPR/GLP-1R</i>	Dual GIP/GLP-1 RA (ultra-long-acting, monthly)	Phase 1	NCT06606106	—	High	Eli Lilly	SC
LY4167586	<i>Undisclosed</i>	Undisclosed	Phase 1	NCT07225556	—	—	Eli Lilly	SC
CT996 / RO7795081	<i>GLP-1R</i>	Oral GLP-1 RA	Phase 1	NCT07081958	—	High	Roche / Carmot	PO
SCO-094	<i>GIPR/GLP-1R</i>	Dual GLP-1/GIP RA	Phase 1	—	—	Moderate	Scovia Pharma	SC
NNC0174-1213	<i>Undisclosed</i>	Undisclosed	Phase 1	NCT06719011	—	—	Novo Nordisk	—
NNC6989-0001	<i>Undisclosed</i>	Undisclosed	Phase 1	NCT07437079	—	—	Novo Nordisk	—
NNC0662-0419	<i>Undisclosed</i>	Undisclosed	Phase 1	NCT07525791	—	—	Novo Nordisk	—
HM15275	<i>GIPR/GLP-1R/GCGR</i>	Next-gen triple agonist	Phase 1	NCT07205900	—	Moderate	Hanmi Pharma	SC
Naperiglipron	<i>GLP-1R</i>	Oral small-molecule GLP-1 RA	Phase 1	NCT07232732	—	High	AstraZeneca	PO
MDR-001	<i>Undisclosed</i>	Undisclosed	Phase 1	NCT06606483	8.9	—	Mediar Therapeutics	—
UBT251	<i>Undisclosed</i>	Undisclosed	Phase 1	NCT07177469	19.7	—	—	—
HDM1002	<i>Undisclosed</i>	Undisclosed	Phase 1/2	NCT06500299	—	—	Harbour BioMed	—
SYH2082	<i>Undisclosed</i>	Undisclosed	Phase 1	NCT07532655	—	—	—	—
Withdrawn								
Rimonabant	<i>CB1R</i>	CB1 inverse agonist	Withdrawn (2008)	RIO-Europe	6.5	High	Sanofi	PO
Lorcaserin	<i>5-HT2C-R</i>	5-HT2C agonist	Withdrawn (2020)	BLOOM (NCT00395135)	5.8	High	Eisai / Arena	PO

WL% = maximum reported weight loss (%); GWAS Ev. = strength of supporting GWAS evidence; SC = subcutaneous; PO = oral; QW = once weekly; QD = once daily; TID = three times daily; QM = once monthly; Q2W = every 2 weeks; IV = intravenous. Data current as of June 2026.