



HAL
open science

Pleiotropic Expression Quantitative Trait Loci Are Enriched in Enhancers and Transcription Factor Binding Sites and Impact More Genes

Aitor González, Pascale Paul

► **To cite this version:**

Aitor González, Pascale Paul. Pleiotropic Expression Quantitative Trait Loci Are Enriched in Enhancers and Transcription Factor Binding Sites and Impact More Genes. Computational and Structural Biotechnology Journal, In press, 10.1016/j.csbj.2024.11.019 . hal-04787752

HAL Id: hal-04787752

<https://hal.science/hal-04787752v1>

Submitted on 18 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Pleiotropic Expression Quantitative Trait Loci Are Enriched in Enhancers and Transcription Factor Binding Sites and Impact More Genes.

Aitor González^{a,*}, Pascale Paul^a

5 ^aAix-Marseille Univ, INSERM U1090, TAGC, 13288 Marseille, France.

*Corresponding author: aitor.gonzalez@univ-amu.fr (A. González).

7 **ABSTRACT**

8

9 Integrating expression quantitative trait loci (eQTL) data with genome-wide association studies
10 (GWAS) enables the discovery of pleiotropic gene regulatory variants that influence a wide range of
11 traits and disease susceptibilities. However, a comprehensive understanding of the distribution of
12 pleiotropic QTLs across the genome and their phenotypic associations remain limited. In this study,
13 we systematically annotated genetic variants associated with both trait variation and gene expression
14 changes, focusing specifically on the unique characteristics of pleiotropic eQTLs. By integrating data
15 from 127 eQTL studies and 417 traits from the IEU Open GWAS Project, we identified 476
16 pleiotropic eQTL variants affecting two or more distinct traits. [Our analysis highlighted 5,345 eQTL](#)
17 [candidates potentially linked to gene expression changes across 293 GWAS traits](#). Notably, the 476
18 pleiotropic eQTLs associated with multiple trait categories were localized within a cumulative 2.5
19 Mbp genomic region. These pleiotropic eQTLs were enriched in enhancer regions and CTCF loops,
20 influencing a larger number of genes in closer genomic proximity. Our findings reveal that pleiotropic
21 eQTLs are concentrated within a small fraction of the genome and exhibit distinct molecular features.
22 [Colocalization results are accessible through an interactive web application and UCSC genome](#)
23 [browser tracks at <https://gwas2eqtl.tagc.univ-amu.fr/gwas2eqtl>, facilitating the exploration of](#)
24 [pleiotropic eQTLs and their roles in gene regulation and disease susceptibility](#).

25

26 **Keywords:** genome-wide associations studies (GWAS), expression quantitative trait loci (eQTL);
27 pleiotropy; colocalization

28 **Introduction**

29 The growing number of genome-wide association studies (GWAS) linking genetic variants to
30 phenotypic or disease traits has led to the constant expansion of comprehensive databases such as
31 the NHGRI-EBI GWAS Catalog [1]. As of 2019, the NHGRI-EBI GWAS Catalog included 5,687
32 studies (GWAS), encompassing 71,673 variant-trait associations from 3,567 publications. The
33 ENCODE consortium allowed to define candidate cis-regulatory elements (CREs), which, combined
34 with evolutionary conservation methods, help in identifying regulatory elements under evolutionary
35 pressures [2]. Large scale GWAS studies have further revealed that many genetic variants exhibit
36 pleiotropy, meaning a single gene can influence multiple, seemingly unrelated phenotypes or traits
37 With the extensive availability of genomic data, it is now possible to investigate pleiotropy at various
38 levels, including the variant level [3–5], the genomic region level [5], the gene level [6] and at the
39 level of co-regulated gene groups [7].

40 Previous studies estimate that the human genome contains between 18,000 and 75,000 pleiotropic
41 variants, 7,757 pleiotropic genes, and cumulated pleiotropic regions ranging 180-1,707 megabases
42 (Mb) in the human genome [3–5]. These estimates arise from different GWAS approaches, which
43 likely contain both false positives and false negatives [3].

44 Expression quantitative trait loci (eQTLs) are genetic variants associated with differences in the
45 expression of one or more genes, thereby shaping gene activity in a cell- and tissue-specific manner
46 [3]. Gaining knowledge on eQTLs can thus allow us to decipher the molecular mechanism that
47 associates these variants to GWAS traits [8]. However, due to linkage disequilibrium, these methods
48 may fail to identify the causal variant. Statistical colocalization of eQTLs and GWAS variants, as
49 well as eQTLs annotation of GWAS traits has been explored [9]. [While ColocDB includes extensive
50 colocalization of eQTLs and GWAS variants, analysis to identify pleiotropic eQTLs is currently
51 missing \[10\].](#) Because many functional common variants lie outside coding sequences, annotating
52 GWAS variants with eQTLs is crucial to uncover the genes and tissues underlying GWAS traits,
53 particularly for identifying the characteristics of pleiotropic eQTLs.

54 Previous research has shown that pleiotropic genomic regions and variants exhibit specific
55 properties at the genomic and cellular levels, regulating a larger number of gene targets and
56 showing activity in a broader range of tissues [5,11,12]. Pleiotropic variants also appear more
57 frequent and exhibit higher effect sizes [3]. In this study, we leveraged two extensive eQTL and
58 GWAS datasets from the EBI eQTL Catalogue [13] and the IEU OpenGWAS Project [14],
59 respectively, to conduct a systematic colocalization analysis based on 127 eQTL studies and 417
60 GWAS across various cell types, tissues, and phenotypes. By categorizing these 417 traits, we
61 identified pleiotropic variants as those associated with two or more categories. This categorization
62 enabled us to develop an online application tool for exploring eQTL and GWAS colocalization,
63 accessible at <https://gwas2eqtl.tagc.univ-amu.fr/gwas2eqtl>, allowing users to investigate
64 colocalization patterns across diverse phenotypes and tissues.

65

66 **Materials and methods**

67 Summary statistics of expression quantitative trait Loci (eQTL) studies

68 We utilized data from 127 eQTL studies available in the EBI eQTL Catalogue, one of the most
69 comprehensive resources providing uniformly processed eQTLs across diverse tissues and cell
70 types [13] (Supplementary Table 1). We downloaded the summary statistics for these studies,
71 representing 127 biological samples, from the EBI eQTL Catalogue. To streamline analyses across
72 heterogeneous cell types and tissues, we categorized these into 35 distinct groups (Supplementary
73 Table 1). Tissue grouping was based on anatomical proximity, such as grouping various brain regions
74 or segments of the digestive tract (e.g., colon). Circulating or immortalized cell types were
75 categorized according to functional roles, resulting in classifications such as blood, immune system,
76 and lymphoblastoid cells.

77

78 Summary Statistics of genome-wide association studies (GWAS)

79 GWAS summary statistics were sourced from the IEU OpenGWAS Project, which aggregates data
80 from several databases, including the EBI's comprehensive GWAS database, manually curated
81 GWAS datasets, and the UK Biobank [14]. We selected GWAS traits from the IEU OpenGWAS
82 Project based on the following criteria: (1) exclusion of molecular traits (e.g., proteomic or
83 methylome data) to focus on colocalization of eQTLs with disease-related variants; (2) inclusion of
84 studies with European ancestry samples, aligning with the population represented in the EBI eQTL
85 Catalogue; (3) stringent definitions of medical or physiological conditions, with environmental or
86 self-reported traits excluded due to limited tissue linkage; and (4) studies with sample sizes of at least
87 10,000 participants, including a minimum of 2,000 cases and 2,000 controls. After applying these
88 filters, 417 GWAS traits remained (see Supplementary Table 2), covering 10,621 clumped lead
89 variants with p-values below 5×10^{-8} from 335 GWAS, which were used to annotate eQTLs.

90

91 Additional genomic datasets, tools

92 This study incorporated additional datasets and tools. Besides eQTL and GWAS datasets, we used
93 transcription factor ChIP-seq peaks and cis-regulatory modules from the ReMap database [15], UCSC
94 annotation data [16], and candidate enhancers ENCODE project's SCREEN database
95 (<https://screen.encodeproject.org/>) [17]. CTCF ChIA-PET loop data were also downloaded from
96 ENCODE as bedpe files. Data exploration was conducted using the UCSC Genome Browser and the
97 OMIM database (<https://omim.org>). and Pipelines were executed using Snakemake, which
98 streamlined and automated the workflows [18].

99

100 Colocalization analysis

101 For the colocalization analysis, complete association data were obtained from OpenGWAS and
102 converted to hg38 coordinates using Picard and Crossmap [19]. Top hit variants from the
103 corresponding GWAS were selected based on a p-value threshold of $5E-8$, using a clumping

104 parameter of $r^2=0.1$ within a 1 Mb clumping window. Within a 1 Mb radius of these top hits,
105 significant eQTL and GWAS variants with a p-value under $5E-8$ were extracted. Missing allele
106 frequencies were filled in using data from the European population in the 1000 Genome Project [20].
107 Variants meeting various criteria, variant frequency strictly between 0 and 1, non-duplicated, and
108 without missing data, were retained. Colocalization between eQTL and GWAS variants was assessed
109 using the "coloc.abf" function from CRAN coloc package. Colocalization between eQTL and GWAS
110 variants was assessed using the coloc.abf function from the CRAN coloc. This analysis was
111 conducted for each window, eQTL sample, and GWAS. The coloc software was specifically utilized
112 to calculate the probabilities that a variant is associated with two distinct traits simultaneously[9]. We
113 prioritized the coloc software to leverage the R interface between our datasets and the tools provided
114 by the IEU OpenGWAS Project. Throughout this paper, analyses were performed with a cutoff of
115 $PP.H4.abf \geq 0.75$ and $SNP.PP.H4 \geq 0.5$, unless explicitly stated otherwise. The parameter $PP.H4.abf$
116 represents the posterior probability that there is a shared causal variant affecting both traits (e.g.,
117 GWAS and eQTL) at a given locus. The parameter $SNP.PP.H4$ is the posterior probability for an
118 individual variant at that locus to be the shared causal variant. We adopted $PP.H4.abf \geq 0.75$ to
119 facilitate comparison with a previous large colocalization analysis using the same software [8].

120

121 Statistical analysis

122 For statistical analyses, we used the Fisher's exact test, the Mann-Whitney U test, the odds ratio test
123 and the Spearman correlation. We set the following significance levels: ns (not significant) ≥ 0.05 ; *
124 < 0.05 ; ** < 0.01 ; *** < 0.001 ; **** < 0.0001 . Details of the statistical tests are provided in the figure
125 legends.

126 Results

127

128 Colocalization of variants associated with GWAS traits and gene expression changes

129 We developed a comprehensive pipeline for colocalization of variants from GWAS and eQTL studies
130 (see Section Methods for details). Variants within the MHC locus (chromosome 6: 25,000,000-
131 35,000,000 bp, hg38) were excluded from this analysis due to their complex linkage disequilibrium
132 structure. The pipeline resulted in colocalization between traits and expression associations for
133 138,274 variants derived from 293 GWAS and 127 eQTL studies, applying a cutoff value of
134 $PP.H4.abf \geq 0.75$ (see Data availability for details). [Following the selection of individual variants at](#)
135 [the locus with moderate probability \(\$SNP.H4.PP \geq 0.5\$ \) of being the shared causal variant, we retained](#)
136 [5,345 variants and 7,040 eQTL genes \(Supplementary Table 4\).](#) The cutoff values of $PP.H4.abf \geq 0.75$
137 and $SNP.H4.PP \geq 0.5$ were selected based on a previous colocalization analysis conducted by Mu et
138 al. [9]. These 5,345 variants were used for subsequent analyses.

139

140 Differential contribution of eQTLs to GWAS variant interpretation across disease categories

141 In our analysis, we aimed to characterize disease variants by leveraging the extensive number of co-
142 localized eQTLs and GWAS variants identified earlier. Specifically, we assessed the proportion of
143 leading GWAS variants that could be elucidated by colocalization with eQTLs within a 1 Mbp
144 window (Supplementary Table 3). Our findings demonstrated significant variability, with between
145 25% and 80% of leading GWAS variants being explained by eQTLs. Notably, the top two
146 categories—sleep disorders and female reproductive system diseases—showed complete explanation
147 by eQTLs; however, these values were derived from only one GWAS with few loci (Fig. 1a,
148 Supplementary Table 3). Overall, the extent to which eQTLs explain GWAS variants varies
149 markedly, ranging from 25% to 100% across different disease categories.

150

151 Web application for exploring pleiotropic variations associated with traits and gene
152 expression

153 To improve accessibility to this valuable resource, we developed a web application for exploring
154 colocalized eQTLs and GWAS traits (<https://gwas2eqtl.tagc.univ-amu.fr/gwas2eqtl/>). This tool
155 allows users to investigate pleiotropic variants associated with a diverse array of traits, including
156 autoimmune diseases, breast cancer, cardiovascular diseases, and height. Additionally, we provide a
157 UCSC track featuring annotated variants (see Data Availability). Supplementary Fig. 1a-c presents
158 three screenshots of these UCSC tracks for B-cells, monocytes, and frontal cortex biological samples,
159 all mapped to chromosome 5 around the IL4 gene.

160 Importantly, the UCSC track illustrates that associated eQTLs and genes vary significantly based on
161 cell type and tissue distribution. For example, in B-cells, identified eQTLs predominantly target the
162 SLC22A5 and MIR3936HG genes (see Supplementary Fig. 1a). In monocytes, eQTLs also target
163 P4HA2 and SLC22A4 genes (see Supplementary Fig. 1b). In contrast, eQTLs in the prefrontal cortex
164 predominantly affect PDLIM4 and RAD50 (see Supplementary Fig. 1c). This variation underscores
165 that the gene targets of eQTLs co-localizing with different GWAS traits exhibit significant differences
166 depending on the specific cell type.

167

168 Coherent clustering of traits based on eQTL effects

169 In the preceding sections, we annotated variants based on associated traits and gene expression. Here,
170 we aim to check if different traits cluster coherently based on the eQTL effects of co-localized
171 eQTL/GWAS variants. To achieve this, we manually categorized 417 GWAS traits into 35 groups,
172 clustering identical or similar traits (see Supplementary Table 2). For each trait, we extracted the beta
173 effect values of various co-localized eQTLs ($PP.H4.abf \geq 0.75$ and $SNP.PP.H4 \geq 0.5$) across different
174 tissues and computed distances between the vectors of different traits using Spearman correlation.

175 To refine the number of traits and categories, we retained only those traits that showed a correlation
176 of at least 0.05 or -0.05 with at least 30 other traits. This analysis revealed coherent clustering among

177 traits related to autoimmune diseases, circulatory system disorders, cancers, and allergies, indicating
178 that similar diseases share common eQTLs (see Fig. 1b). Notably, skin cancer exhibited overlapping
179 signals with autoimmune diseases, such as type 1 diabetes and digestive autoimmune disorders (see
180 Fig. 1b), suggesting potential comorbidities.

181 182 Identification of eQTLs associated with two or more trait categories.

183 In the previous section, we categorized traits into distinct groups and postulated that these categories
184 could be utilized to identify pleiotropic eQTLs associated with multiple trait categories. We annotated
185 eQTLs according to these trait categories, categorizing them based on the number of categories they
186 encompass. Using a cutoff of $\text{SNP.PP.H4} \geq 0.5$, we identified 476 pleiotropic eQTLs linked to two
187 or more categories (refer to Supplementary Table 4). Notably, the most pleiotropic eQTLs were
188 associated with categories covering a range of traits, including autoimmune diseases, circulatory
189 diseases, and cancer (see Table 1). These eQTLs are located near genes such as ALDH2 (12q24.12),
190 ENO1 (1p36.23), ORMDL3 (17.q21.1), IL4 (5q31.1), BACH2 (6q15), FUT2 (19q13.33), SLC39A8
191 (4q24), and CYP1A1 (15q24.1) (see Table 1). Supplementary Figure 1 illustrates the genomic region
192 around the IL4 gene, highlighting highly pleiotropic eQTLs such as rs2522051 and rs17622656 along
193 with their target genes.

194 In a comparative analysis with another study on pleiotropy, where trait categories were aggregated
195 based on genomic proximity, we found that the eQTLs identified in our study were associated with
196 significantly more categories (see Supplementary Fig. 2a) [5]. Further examination demonstrated that
197 for eQTLs associated with one, two, and more categories, 10-15% of our eQTLs overlapped with
198 those from the previous study (refer to Supplementary Fig. 2b). In conclusion, we identified 476
199 pleiotropic eQTLs associated with a diverse array of diseases, including autoimmune diseases,
200 cardiovascular disorders, hypertensive disorders, mental or behavioral conditions, and
201 musculoskeletal disorders.

202

203 Specificity of eQTLs across traits, gene targets, and tissues

204 In our analysis, we classified eQTLs based on the number of trait categories and explored the
205 relationship between this count and the number of target genes and tissues. We aimed to determine
206 whether eQTLs exhibit specificity to particular traits, gene targets, and tissues. We computed the
207 proportion of eQTLs annotated with various counts of trait categories, gene targets, and tissues.

208 The results indicate that the majority of eQTLs (approximately 90%) are associated with a single trait
209 category, while about 10% are linked to two trait categories, and less than 1% to three or more (see
210 Fig. 2a). Repeating this analysis for eQTLs of each trait category revealed that around 80% of eQTLs
211 related to allergic diseases are involved in one category, compared to only 60% for those in
212 cardiovascular diseases, which are associated with two or more categories (see Supplementary Fig.
213 3a and 3b). This suggests that eQTLs in cardiovascular-related diseases tend to be more pleiotropic.

214 Regarding gene targets, approximately 50% of eQTLs modify the expression of a single gene, 20%
215 affect two genes, and the remaining eQTLs influence three or more genes (see Fig. 2b). Concerning
216 tissues, around 40% of eQTLs are found in a single tissue, approximately 20% in two tissues, and the
217 rest in three or more (see Fig. 2c).

218 We further explored the correlation between the number of GWAS categories, genes, and tissues.
219 The counts of eQTL genes and tissues showed a high correlation ($r=0.79$), while the count of trait
220 categories exhibited a weaker correlation with both gene and tissue counts (see Fig. 2d). In summary,
221 most eQTLs are associated with a single trait category, although these eQTLs regulate multiple genes
222 across various tissues.

223

224 Concentration of pleiotropic eQTLs in 2.5 Mbp cumulated genomic regions

225 Our investigation into pleiotropic eQTLs revealed that these loci are concentrated in specific regions
226 of the genome. Notably, cytobands such as 3q23, 5q31.1, 9p21.3, 15q24.1, and 19q13.33 emerge as
227 key locations for the most pleiotropic eQTLs (see Table 1 and Supplementary Table 4). This

228 observation led us to explore whether a small fraction of the genome contains the majority of
229 pleiotropic eQTLs.

230 To identify genomic regions associated with pleiotropic eQTLs, we used a sliding window of 100,000
231 bp, aggregating regions containing pairs of co-localized eQTL variants annotated with two or more
232 categories, located within a maximal distance of 100,000 bp. Each eQTL was assigned to a single
233 region. The complete list of regions with annotations is provided in Table 2 and Supplementary Table
234 5. Our findings indicate that regions with five or more trait categories encompass approximately 0.8
235 Mbp of the genome, while regions with four trait categories total 1.5 Mbp, and those with three or
236 more trait categories account for 2.5 Mbp (see Fig. 2e). Additionally, regions containing at least two
237 trait categories and measuring less than 10,000 bp constitute around 60% of the total pleiotropic
238 regions, whereas regions up to 100,000 bp comprise 90% of the total pleiotropic regions (see Fig. 2f).
239 The most pleiotropic region, associated with nine trait categories, is located at 5:131,912,097-
240 132,802,472 in cytoband 5q31.1, encompassing genes such as the interferon response factor 1 (IRF1)
241 and interleukins IL3, IL4, IL5, and IL13 (refer to Table 2). The largest region, spanning 7:2,712,518-
242 7,254,268 in cytoband 7p22.3, measures 4,541,751 bp and contains variants linked to autoimmune
243 diseases, respiratory diseases, and height (see Supplementary Table 5). This analysis underscores that
244 pleiotropic eQTLs are highly concentrated in a small fraction of the genome.

245

246 Pleiotropic eQTLs exhibit lower effects on traits, increased significance, and higher variant
247 frequencies

248 Building on previous studies indicating that pleiotropic variants tend to have lower effects and higher
249 frequencies [3], our analysis explored how the effects (beta), association significance, and variant
250 frequencies vary based on the number of trait categories associated with pleiotropic eQTLs.

251 We analyzed the average absolute eQTL and GWAS effect sizes (beta) in relation to the increasing
252 number of trait categories. We observed that the effect size on gene expression is significantly
253 different, although the direction of the changes is inconsistent across varying levels of pleiotropy (see

254 Fig. 3a). In contrast, the effect size on the GWAS trait clearly decreases with increasing eQTL
255 pleiotropy (see Fig. 3b). Further examination of the significance of eQTL and GWAS effects revealed
256 a notable increase for both gene expression and GWAS traits (see Fig. 3c, d).

257 Comparing average effect sizes from different studies can be challenging, as these depend on study
258 size and variant frequencies. To homogenize these variables, we repeated the comparison of eQTL
259 and GWAS effects using variants with restricted frequencies (0.45-0.55) or sample sizes (75,000-
260 125,000) (see Supplementary Fig. 4). These results confirmed the previous observations that GWAS
261 effects decrease in pleiotropic eQTLs.

262 We also explored the relationship between variant frequency and pleiotropy. We partition variant
263 frequencies of each population by the trait category count. The variant frequencies of pleiotropic
264 eQTLs were significantly higher in the European population (see Fig. 3e). In other populations, we
265 did not observe a significantly higher variant frequency, potentially influenced by the fact that
266 colocalization was computed in the European population (Data not shown). In summary, our analysis
267 suggests that pleiotropic eQTLs exhibit lower effects on traits, are more significant, and are more
268 frequent in the European population.

269

270 Pleiotropic eQTLs are more likely missense variants.

271 To gain insights into the variant effect consequences of pleiotropic eQTLs compared to non-
272 pleiotropic eQTLs, we utilized the EBI variant effect predictor (VEP) for annotation [21].
273 Subsequently, we calculated odds ratios and performed Fisher exact tests for pleiotropic eQTLs. The
274 analysis revealed that pleiotropic eQTLs exhibit significantly higher odds ratios for being missense
275 variants, with values of two and nine for category counts of two and more, respectively. This finding
276 strongly suggests that pleiotropic eQTLs are more likely to be associated with missense variants.

277

278 Pleiotropic eQTLs are closer to gene targets

279 In our exploration of pleiotropic eQTLs, we delved into understanding their spatial relationship with
280 gene targets, focusing on the distance to the closest and the most distal gene. The analysis revealed
281 that the median distance of eQTLs associated with one, two, and more than two trait categories to
282 their closest gene is 35 kbp, 23 kbp, and 18 kbp, respectively (see Fig. 4a). This suggests a significant
283 reduction in the distance of pleiotropic eQTLs to their closest gene targets. However, when analyzing
284 the distance of pleiotropic eQTLs to the furthest gene target, no significant difference was found
285 (Data not shown). This indicates that the distance of the eQTL to its furthest gene target remains
286 relatively consistent in pleiotropic eQTLs. In summary, our analysis suggests that pleiotropic eQTLs
287 exhibit a shorter distance to their closest gene targets, potentially reflecting a more direct influence
288 on gene regulation.

289

290 Pleiotropic eQTLs have lower gene specificity

291 In our investigation into the characteristics of pleiotropic eQTLs, we turned our attention to their
292 relationship with the number of target genes. Our analysis involved computing the cumulative
293 proportion of eQTLs associated with different numbers of target genes in specific tissues. The
294 findings revealed that the proportion of eQTLs targeting only one gene was approximately 0.72, 0.65,
295 and 0.6 for eQTLs associated with one, two, and more than two trait categories, respectively (see Fig.
296 4b). This trend was similarly observed for eQTLs targeting two genes. This observation could be
297 related to the higher variant frequency of pleiotropic eQTLs. [To homogenize variant frequencies, we](#)
298 [repeated this analysis with variants that show a frequency in the European population between 0.45](#)
299 [and 0.55 \(Supplementary Fig. 5b\)](#). Both analyses suggest that pleiotropic eQTLs tend to modulate a
300 larger number of genes, indicating a broader impact on gene regulation compared to non-pleiotropic
301 eQTLs.

302

303 Cumulative proportion analysis reveals lower tissue specificity of pleiotropic eQTLs

304 Continuing our exploration of pleiotropic eQTLs, we turned our focus to their relationship with the
305 number of tissue categories. Our analysis involved computing the cumulative proportion of eQTLs
306 targeting genes in each number of tissues.

307 The observations revealed systematic higher proportions of non-pleiotropic eQTLs in a low number
308 of tissues (Supplementary Fig. 5a). This observation could be related to the higher variant frequency
309 of pleiotropic eQTLs. To homogenize variant frequencies, we repeated this analysis with variants that
310 show a frequency in the European population between 0.45 and 0.55 (Supplementary Fig. 5c). Both
311 analyses suggest that pleiotropic eQTLs exhibit a tendency to be less tissue-specific, influencing gene
312 expression across a broader range of tissues compared to their non-pleiotropic counterparts.

313

314 Enhancer Properties of Pleiotropic eQTLs

315 In our quest to understand the characteristics of pleiotropic eQTLs, we explored their association with
316 enhancer regions and the binding of transcription factors (TFs). Utilizing the ReMap database [15],
317 we examined the number of unique transcription factors bound within a radius of 10 bp around each
318 eQTL, using a window of 20 bp. We show that pleiotropic eQTLs associated to two and three or more
319 trait categories are significantly enriched in ENCODE enhancer regions with odds ratios 1.5 and 2.5,
320 respectively (Fig. 4c). We also find that pleiotropic eQTLs associated to two and three or more trait
321 categories are significantly enriched in peaks of ReMap transcription factors with odds ratios 1.5 and
322 2.5, respectively (Fig. 4d). These two results suggest that gene regulatory regions show more robust
323 properties for pleiotropic eQTLs. We show that eQTLs associated with one, two, or more trait
324 categories exhibit a median binding of 7, 10, and 9 transcription factors, respectively (Fig. 4e). This
325 suggests that pleiotropic eQTLs tend to be associated with a higher number of TFs.

326 Enhancers are likely enclosed within CTCF loops [22]. To investigate the relationship of pleiotropic
327 eQTLs/gene links and CTCF loops, we evaluated the enrichment of the regions linking pleiotropic
328 eQTLs and target genes withing CTCF ChIA-PET loops in several cell lines from ENCODE. Our

329 results shows that the links between pleiotropic eQTLs and target genes are enriched in the CTCF
330 loops of cell lines A549 (Adenocarcinomic alveolar basal epithelial cells), CD8-T-cells, GM10248
331 (Lymphoblastoid cell line), HUVEC (Human umbilical vein endothelial cells), K562
332 (Erythroleukemia) and WTC11 (Human induced pluripotent stem cell) (Fig. 4f and Supplementary
333 Fig. 6).

334 We investigated the presence of cis-regulatory modules (CRMs), which are non-coding genomic
335 regions with a higher density of bound transcription factors [15]. The odds ratio of variants annotated
336 with CRMs versus non-annotated was 1, 1.5, and 1.9 for eQTLs associated with one, two, or more
337 trait categories, respectively (see Fig. 4d). This implies that pleiotropic eQTLs are not only bound by
338 more transcription factors but are also more likely to be in cis-regulatory modules.

339 Discussion

340 Genome-wide association studies have uncovered numerous loci linked to various diseases, however
341 the specific causal contributions of pleiotropic variants within these loci remain largely elusive. This
342 study leveraged on two extensive databases, the IEU OpenGWAS project and the EBI eQTL
343 Catalogue, encompassing summary statistics from 417 GWAS and 127 eQTL datasets, to conduct a
344 comprehensive colocalization pipeline. Our objective was to clarify the characteristics of pleiotropic
345 eQTLs, contributing to a more precise understanding of these variants.

346 These findings are accessible via an interactive web platform [https://gwas2eqtl.tagc.univ-](https://gwas2eqtl.tagc.univ-amu.fr/gwas2eqtl/)
347 [amu.fr/gwas2eqtl/](https://gwas2eqtl.tagc.univ-amu.fr/gwas2eqtl/), enabling users to explore these colocalizations and analyze the dataset, thus
348 providing a valuable resource. Recently, COLOCdb, a database of colocalization data across 3,000
349 GWAS and 13 xQTL types, has been published but this database was not used to systematically
350 investigate the properties of human genetic variants [10].

351 Our study revealed that approximately 50% of the prominent variants associated with autoimmune
352 diseases can be elucidated by eQTLs. This percentage surpasses previous studies reporting 25–38%
353 explanatory power using eQTLs and splicing QTLs (sQTLs) [8,23]. This explanatory capacity may
354 stem from our broader inclusion of 127 eQTL datasets. In contrast, prior studies, such as Mu et al.

355 [8], used a more limited immune cell types fore eQTLs and sQTLs potentially missing contributions
356 in other tissues. Similarly, Connally et al. focused on breast tissue-sepecific eQTLs to study breast
357 cancer, possibly overlooking immune cell interactions involved in cancer biology [24]. Our study
358 suggests that colocalization studies may lack the breadth of eQTL coverage necessary to to fully
359 account for GWAS findings across diverse traits [25].

360 Our colocalization results also demonstrate remarkable consistency across independent datasets. For
361 instance, related autoimmune diseases affecting the digestive system, such as inflammatory bowel
362 disease, Crohn’s disease, and ulcerative colitis, showed strong colocalization coherence, despite
363 originating from distinct GWAS sources [25–27]. Our analysis extended to predict overlaps with
364 other autoimmune diseases, including type 1 diabetes, and traits associated with digestive
365 autoimmune and inflammatory diseases, and skin cancer. These findings underscore the potential of
366 pleiotropic eQTLs in drug repurposing [28].

367 Additionally, our data highlights the stability of cell-specific eQTL patterns across different
368 independent datasets. Although the effects of quantitative trait loci (QTLs) vary widely between cell
369 types, specific cell types maintain consistent patterns across studies. For example, the colocalization
370 results for eQTLs targeting the gene SLC22A5 in B-cells exhibited similar results across three
371 different studies [29–31]. This consistency is was also observed in monocytes and cortex samples.

372 Our analysis of pleiotropic eQTLs across varying SNP.PP.H4 cutoff values (0.25, 0.5, and 0.75)
373 confirmed stable identification of key variants. High-pleiotropy variants, such as rs11065979,
374 rs11072508, rs13107325, rs301802, rs2522051, rs601338, and rs72928038, were consistently
375 detected at all three cutoff values. An intermediate cutoff value of 0.5 was chosen to strike a balance
376 between sensitivity and specificity in identifying colocalized eQTLs.

377 The functional relevance of identified pleiotropic variants aligns with current literature. For instance,
378 rs13107325, a pleiotropic eQTL associated with the SLC39A8 gene, has been validated in knock-in
379 mouse models [32]. Similarly, the physiological impact of rs72928038 on the BACH2 gene was
380 recently confirmed in b-cell lines and mutant mice [33].

381 Among the most pleiotropic regions, such as those around ABO, CDKN2A, CHRNA5, CYP1A1,
382 EEF1A2, FURIN, IL4, MC1R, and TP53, the selection of regions with pleiotropic eQTLs remained
383 insensitive to changes in the SNP.PP.H4 cutoff value. Notably, genes like ABO, CDKN2A, and
384 FURIN, known for their pleiotropic effects on a wide array of phenotypes, were consistently
385 identified [33–35].

386 In alignment with prior reports, we observed that pleiotropic eQTLs tend to be less tissue-specific
387 [3,5]. Additionally, our findings indicate that pleiotropic eQTLs are less gene-specific, aligning with
388 previous studies demonstrating that eQTLs affecting multiple neighboring genes exhibit greater
389 pleiotropy [12]. Mechanistically, this aligns with the notion that eQTLs influencing a higher number
390 of genes and active in more tissues are associated with a broader range of traits. Pleiotropic variants
391 were also found to be more frequent, a trend that might be a statistical artifact, where more frequent
392 variants are seemingly more likely to be associated with specific diseases [3]. Consistent with
393 previous studies, we noted that pleiotropic eQTLs are enriched in proximity to genes [5].

394 The differences between pleiotropic and non-pleiotropic eQTLs become evident at short distances.
395 For example, the eQTL rs2522051 in the IL4 locus is associated with four trait categories, while
396 nearby non-pleiotropic eQTLs (rs6894249, rs6894249, and rs80112473) at 31 bp, 656 bp, and 1,795
397 bp distances are associated with fewer genes and tissues. The contrast is further illustrated by
398 examples like rs159963 in the ENO1 locus and rs151174 in the SULT1A1 locus, where pleiotropic
399 eQTLs exhibit more gene targets and tissue activity than their non-pleiotropic counterparts.
400 Conversely, some pleiotropic eQTLs, such as rs823118 in the RAB29 locus, rs12656497 in the NPR3
401 locus, rs10051765 in the FRFR14 locus, region chr11:1,865,076-1,877,434 in the TNNT3 locus,
402 rs2071382 in the FES locus, rs8067378 in the ORMDL3 locus, and rs601338 in the FUT2 locus.
403 target fewer genes and tissues than non-pleiotropic counterparts, suggesting additional factors
404 influencing eQTL pleiotropy remain to be identified.

405 [Our study also shows that pleiotropic eQTLs are enriched in enhancer regions and transcription factor](#)
406 [binding sites. This supports previous research indicating that pleiotropic enhancers, active across](#)

407 multiple tissues, comprise a small portion of the genome yet are enriched in transcription factor
408 binding and gene regulatory connections [36].
409 Recent CRISPRi screens highlight that enhancer-target gene interactions are enriched within
410 topological associated domains and CTCF ChIA-PET loops [22,37], which could serve to get genes
411 from GWAS variants within specific contexts [38]. In our findings, pleiotropic eQTLs and their target
412 genes also exhibited enrichment in CTCF loops.
413 Our analysis reveals that pleiotropic eQTLs generally show reduced GWAS effect sizes (beta),
414 greater significance in both eQTL and GWAS associations, and higher allele frequencies.
415 Furthermore, pleiotropic eQTLs enriched in enhancers and transcription factor binding sites regulate
416 a broader set of genes within CTCF loops

417

418 **Authors' contributions**

419 Aitor González: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources,
420 Data Curation, Visualization, Writing - original draft, Writing - review & editing.

421 Pascale Paul: Conceptualization, Software, Writing - review & editing.

422 **Competing interests**

423 The authors assert that they have no competing interests.

424 **Acknowledgements**

425 We express our gratitude to the Centre de Calcul Intensif d'Aix-Marseille for generously providing
426 access to its high-performance computing resources. Special thanks to Léopoldine Lecerf for her
427 contribution to the initial version of the pipeline developed during her internship.

428 **Code availability**

429 The code for the GWAS/eQTL colocalization pipeline is accessible on GitHub at
430 <https://github.com/aitgon/gwas2eqtl>. Additionally, the code to replicate the pleiotropy results can be
431 found at https://github.com/aitgon/gwas2eqtl_pleiotropy.

432 **Data availability**

433 An interactive table for searching and browsing GWAS/eQTL colocalizations with $PP.H4.abf \geq 0.75$
434 is accessible through the website application: <http://gwas2eqtl.tagc.univ-amu.fr/gwas2eqtl>. Genome
435 browser tracks displaying GWAS/eQTL colocalizations with $PP.H4.abf \geq 0.75$ are accessible at
436 <https://genome-euro.ucsc.edu/s/agonzalez/gwas2eqtl>. The raw table of colocalizations at $PP.H4.abf \geq$
437 0.75 and the full colocalization data can be downloaded from Zenodo:
438 <https://doi.org/10.5281/zenodo.14064652>.

439

440 **Declaration of generative AI and AI-assisted technologies in** 441 **the writing process**

442 During the preparation of this work the authors used <https://chatgpt.com/> in order to improve language
443 and readability. After using this tool/service, the authors reviewed and edited the content as needed
444 and take full responsibility for the content of the publication.

445

446 **References**

- 447 [1] Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The
448 NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and
449 summary statistics 2019. *Nucleic Acids Res* 2019;47:D1005–12.
450 <https://doi.org/10.1093/nar/gky1120>.
451 [2] Andrews G, Fan K, Pratt HE, Phalke N, ZOONOMIA CONSORTIUM, Karlsson EK, et al.
452 Mammalian evolution of human cis-regulatory elements and transcription factor binding sites.
453 *Science* 2023;380:eabn7930.

- 454 [3] Shikov AE, Skitchenko RK, Predeus AV, Barbitoff YA. Phenome-wide functional dissection of
455 pleiotropic effects highlights key molecular pathways for human complex traits. *Sci Rep*
456 2020;10:1037. <https://doi.org/10.1038/s41598-020-58040-4>.
- 457 [4] Jordan DM, Verbanck M, Do R. HOPS: a quantitative score reveals pervasive horizontal
458 pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and
459 diseases. *Genome Biol* 2019;20:222. <https://doi.org/10.1186/s13059-019-1844-7>.
- 460 [5] Watanabe K, Stringer S, Frei O, Umićević Mirkov M, de Leeuw C, Polderman TJC, et al. A
461 global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*
462 2019;51:1339–48. <https://doi.org/10.1038/s41588-019-0481-0>.
- 463 [6] Chesmore K, Bartlett J, Williams SM. The ubiquity of pleiotropy in human disease. *Hum Genet*
464 2018;137:39–44. <https://doi.org/10.1007/s00439-017-1854-z>.
- 465 [7] Barrio-Hernandez I, Schwartzentruber J, Shrivastava A, Del-Toro N, Gonzalez A, Zhang Q, et
466 al. Network expansion of genetic associations defines a pleiotropy map of human cell biology.
467 *Nat Genet* 2023;55:389–98. <https://doi.org/10.1038/s41588-023-01327-9>.
- 468 [8] Mu Z, Wei W, Fair B, Miao J, Zhu P, Li YI. The impact of cell type and context-dependent
469 regulatory variants on human immune traits. *Genome Biol* 2021;22:122.
470 <https://doi.org/10.1186/s13059-021-02334-x>.
- 471 [9] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian
472 test for colocalisation between pairs of genetic association studies using summary statistics.
473 *PLoS Genet* 2014;10:e1004383. <https://doi.org/10.1371/journal.pgen.1004383>.
- 474 [10] Pan S, Kang H, Liu X, Li S, Yang P, Wu M, et al. COLOCdb: a comprehensive resource for
475 multi-model colocalization of complex traits. *Nucleic Acids Research* 2024;52:D871–81.
476 <https://doi.org/10.1093/nar/gkad939>.
- 477 [11] GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human
478 tissues. *Science* 2020;369:1318–30. <https://doi.org/10.1126/science.aaz1776>.
- 479 [12] Ribeiro DM, Rubinacci S, Ramisch A, Hofmeister RJ, Dermitzakis ET, Delaneau O. The
480 molecular basis, genetic control and pleiotropic effects of local gene co-expression. *Nat*
481 *Commun* 2021;12:4842. <https://doi.org/10.1038/s41467-021-25129-x>.
- 482 [13] Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, et al. A compendium
483 of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet*
484 2021;53:1290–9. <https://doi.org/10.1038/s41588-021-00924-w>.
- 485 [14] Lyon MS, Andrews SJ, Elsworth B, Gaunt TR, Hemani G, Marcora E. The variant call format
486 provides efficient and robust storage of GWAS summary statistics. *Genome Biol* 2021;22:32.
487 <https://doi.org/10.1186/s13059-020-02248-0>.
- 488 [15] Hammal F, de Langen P, Bergon A, Lopez F, Ballester B. ReMap 2022: a database of Human,
489 Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-
490 binding sequencing experiments. *Nucleic Acids Res* 2022;50:D316–25.
491 <https://doi.org/10.1093/nar/gkab996>.
- 492 [16] Nassar LR, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, et al. The UCSC
493 Genome Browser database: 2023 update. *Nucleic Acids Res* 2023;51:D1188–95.
494 <https://doi.org/10.1093/nar/gkac1072>.
- 495 [17] The ENCODE Project Consortium, Abascal F, Acosta R, Addleman NJ, Adrian J, Afzal V, et
496 al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*
497 2020;583:699–710. <https://doi.org/10.1038/s41586-020-2493-4>.
- 498 [18] Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, et al. Sustainable
499 data analysis with Snakemake. *F1000Res* 2021;10:33.
500 <https://doi.org/10.12688/f1000research.29032.2>.
- 501 [19] Zhao H, Sun Z, Wang J, Huang H, Kocher J-P, Wang L. CrossMap: a versatile tool for
502 coordinate conversion between genome assemblies. *Bioinformatics* 2014;30:1006–7.
503 <https://doi.org/10.1093/bioinformatics/btt730>.

- 504 [20] 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM,
505 et al. A global reference for human genetic variation. *Nature* 2015;526:68–74.
506 <https://doi.org/10.1038/nature15393>.
- 507 [21] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant
508 Effect Predictor. *Genome Biol* 2016;17:122. <https://doi.org/10.1186/s13059-016-0974-4>.
- 509 [22] Luo R, Yan J, Oh JW, Xi W, Shigaki D, Wong W, et al. Dynamic network-guided CRISPRi
510 screen identifies CTCF-loop-constrained nonlinear enhancer gene regulatory activity during cell
511 state transitions. *Nat Genet* 2023;55:1336–46. <https://doi.org/10.1038/s41588-023-01450-7>.
- 512 [23] Chun S, Casparino A, Patsopoulos NA, Croteau-Chonka DC, Raby BA, De Jager PL, et al.
513 Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-
514 associated loci in three major immune-cell types. *Nat Genet* 2017;49:600–5.
515 <https://doi.org/10.1038/ng.3795>.
- 516 [24] Connally NJ, Nazeen S, Lee D, Shi H, Stamatoyannopoulos J, Chun S, et al. The missing link
517 between genetic association and regulatory function. *Elife* 2022;11:e74970.
518 <https://doi.org/10.7554/eLife.74970>.
- 519 [25] Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe
520 interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*
521 2012;491:119–24. <https://doi.org/10.1038/nature11582>.
- 522 [26] de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide
523 association study implicates immune activation of multiple integrin genes in inflammatory
524 bowel disease. *Nat Genet* 2017;49:256–61. <https://doi.org/10.1038/ng.3760>.
- 525 [27] Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses
526 identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk
527 across populations. *Nat Genet* 2015;47:979–86. <https://doi.org/10.1038/ng.3359>.
- 528 [28] Reay WR, Cairns MJ. Advancing the use of genome-wide association studies for drug
529 repurposing. *Nat Rev Genet* 2021;22:658–71. <https://doi.org/10.1038/s41576-021-00387-z>.
- 530 [29] Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo
531 J, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*
532 2018;175:1701–1715.e16. <https://doi.org/10.1016/j.cell.2018.10.022>.
- 533 [30] Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, et al. Genetics of gene
534 expression in primary immune cells identifies cell type-specific master regulators and roles of
535 HLA alleles. *Nat Genet* 2012;44:502–10. <https://doi.org/10.1038/ng.2205>.
- 536 [31] Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charlotiaux B, et al. IBD
537 risk loci are enriched in multigenic regulatory modules encompassing putative causative genes.
538 *Nat Commun* 2018;9:2427. <https://doi.org/10.1038/s41467-018-04365-8>.
- 539 [32] Mouri K, Guo MH, de Boer CG, Lissner MM, Harten IA, Newby GA, et al. Prioritization of
540 autoimmune disease-associated genetic variants that perturb regulatory element activity in T
541 cells. *Nat Genet* 2022;54:603–12. <https://doi.org/10.1038/s41588-022-01056-5>.
- 542 [33] Schuch JB, Polina ER, Rovaris DL, Kappel DB, Mota NR, Cupertino RB, et al. Pleiotropic
543 effects of Chr15q25 nicotinic gene cluster and the relationship between smoking, cognition and
544 ADHD. *J Psychiatr Res* 2016;80:73–8. <https://doi.org/10.1016/j.jpsychires.2016.06.002>.
- 545 [34] Liu H, Sun Y, Zhang X, Li S, Hu D, Xiao L, et al. Integrated Analysis of Summary Statistics to
546 Identify Pleiotropic Genes and Pathways for the Comorbidity of Schizophrenia and
547 Cardiometabolic Disease. *Front Psychiatry* 2020;11:256.
548 <https://doi.org/10.3389/fpsy.2020.00256>.
- 549 [35] Arguinano A-AA, Ndiaye NC, Masson C, Visvikis-Siest S. Pleiotropy of ABO gene: correlation
550 of rs644234 with E-selectin and lipid levels. *Clin Chem Lab Med* 2018;56:748–54.
551 <https://doi.org/10.1515/cclm-2017-0347>.
- 552 [36] Singh D, Yi SV. Enhancer Pleiotropy, Gene Expression, and the Architecture of Human
553 Enhancer–Gene Interactions. *Molecular Biology and Evolution* 2021;38:3898–909.
554 <https://doi.org/10.1093/molbev/msab085>.

- 555 [37] Yao D, Tycko J, Oh JW, Bounds LR, Gosai SJ, Lataniotis L, et al. Multicenter integrated
556 analysis of noncoding CRISPRi screens. *Nat Methods* 2024;21:723–34.
557 <https://doi.org/10.1038/s41592-024-02216-7>.
558 [38] Gschwind AR, Mualim KS, Karbalayghareh A, Sheth MU, Dey KK, Jagoda E, et al. An
559 encyclopedia of enhancer-gene regulatory interactions in the human genome 2023.
560 <https://doi.org/10.1101/2023.11.09.563812>.
561

562

563 **Figure and table legends**

564

565 **Figure 1: Proportion of explained GWAS leading loci and trait clustering based on eQTLs.**

566 a) The average proportion of explained loci in GWAS across different trait categories is illustrated.
567 Detailed proportions can be found in Supplementary Table 3, with error bars representing the 95%
568 confidence level.

569 b) The distance between traits is determined based on the Spearman correlation between eQTL
570 coefficients. Only GWAS traits with a minimal number of eQTLs between them were selected. Labels
571 include the dataset source (ebi-a: Datasets meeting minimum requirements imported from the EBI
572 database of complete GWAS summary data; ieu-a: GWAS summary datasets generated by various
573 consortia, initially developed for MR-Base; ukb-a: Neale lab analysis of UK Biobank phenotypes,
574 round 1; ukb-b: IEU analysis of UK Biobank phenotypes; ukb-d: Neale lab analysis of UK Biobank
575 phenotypes, round 2), the PubMed identifier (if available, otherwise 0), and the trait ontology term.

576

577 **Figure 2: Distribution of eQTLs and regions based on trait counts, eQTL gene specificity and** 578 **biological samples.**

579 a-c) The proportion of eQTLs is examined based on the count of trait categories (a), eQTL genes (b),
580 and eQTL biological samples (c). The count of eQTLs is shown over the bars in Figure (a).

581 d) The Spearman correlation between the count of trait categories, eQTL genes, and eQTL biological
582 samples is explored for colocalized eQTLs and GWAS variants. “cnt” stands for count.

583 e, f) To merge colocalized variants to pleiotropic regions, we iteratively merge colocalized variants
584 located at less than 100,000 bp into regions and annotate the regions with the number of trait
585 categories.

586 e) This plot shows the cumulative length of the regions containing colocalized variants starting with
587 the most pleiotropic categories. The number above the bar shows the number of pleiotropic eQTLs.

588 f) This plot shows the cumulative proportion of genomic regions containing eQTLs associated to two
589 or more trait categories with different lengths. Around 60% of these regions show a length of 10,000
590 bp or less.

591

592 **Figure 3: Properties of eQTLs across various pleiotropy levels: insights into variant effects on**
593 **gene expression, trait associations, and allele frequencies**

594 a, b) Average of the absolute effect of variants on gene expression (a) and traits (b).

595 c, d) Average of the negative log₁₀ of the p-value of the association with gene expression (c) and
596 traits (d).

597 e) Average variant frequency in the European population from the 1000 Genomes database. This plot
598 gives the distribution of variant frequencies of colocalized eQTL/GWAS variants partitioned by the
599 trait category count.

600 g) Odds ratio of the eQTLs annotated with the given ENSEMBL variant effect predictor compared to
601 eQTLs with a single trait category count.

602 Median values are shown in the boxes of figures a-d. Boxes of the boxplots show the quartiles of the
603 dataset while the whiskers extend to show the rest of the distribution except outliers. In figures a-e,
604 we carried out the Mann-Whitney U test with a two-sided research hypothesis, and in figure f, the
605 Fisher exact test with a two-sided research hypothesis (Significance: ns ≥ 0.05 * < 0.05 , ** < 0.01 , ***
606 < 0.001 , **** < 0.0001).

607

608 **Figure 4: Exploring eQTL gene distance and biological sample counts, transcription factor**
609 **binding, and CRM annotation across diverse pleiotropy levels**

- 610 a) Closest gene distance for each eQTL in relation to the count of trait categories.
- 611 b) Cumulative proportion of eQTL-tissue pairs for increasing counts of eQTL target genes partitioned
612 by trait category counts.
- 613 c) Odds ratio of eQTLs annotated with ENCODE enhancer versus non-annotated regions for different
614 trait category counts.
- 615 d) Odds ratio of eQTLs annotated with ReMap non-redundant transcription factor peaks versus non-
616 annotated regions for different trait category counts.
- 617 e) Count of bound transcription factors (TF) in the region surrounding eQTLs with a radius of 10 bp
618 (Window 20 bp) for different trait category counts.
- 619 f) Odds ratio of eQTLs within CTCF ChIA-PET loops in isogenic replicate r1 of CD8 T-cells for
620 different trait category counts.
- 621 Boxes of the boxplots depict the quartiles of the dataset, while the whiskers extend to show the rest
622 of the distribution except outliers. In figures a and c, we conducted the Mann-Whitney U test, with a
623 two-sided research hypothesis, and in figure d, the Fisher exact test with a one-sided research
624 hypothesis (Significance: ns ≥ 0.05 * < 0.05 , ** < 0.01 , *** < 0.001 , **** < 0.0001).

625

626 **Table 1: Pleiotropic eQTLs in various cytobands associated with four distinct trait categories.**

627 The term "gene marker" denotes the most frequently cited gene within the eQTL target genes.

628

629 **Table 2: Regions exhibiting pleiotropy involving six or more trait categories.**

630 These regions were constructed around pleiotropic eQTLs linked to two or more trait categories using
631 a sliding window of 100,000 base pairs. The "gene marker" column indicates the most cited eQTL
632 gene in the region, as per NCBI PubMed. The "eQTLs cnt." column provides the count of eQTLs in
633 that region. Genomic coordinates are presented for the hg38 assembly.

634

635 **Supplementary figure and table legends**

636 **Supplementary Figure 1: UCSC tracks screenshot for eQTLs colocalizing with GWAS variants** 637 **in the specified regions.**

638 screenshot for eQTLs colocalizing with GWAS variants in the specified region (chr5:132,137,990-
639 132,728,110, hg38) for B-cells (a), classical monocytes (b), and cortex cells (c):

640 In panel (a), denoting B-cells, the image illustrates different components. The red color signifies
641 positive eQTL beta coefficients, while blue represents negative eQTL beta coefficients. Two sections
642 are provided for each biological sample: "beta equal" corresponds to eQTLs with the same effect sign
643 as the GWAS, and "beta unequal" corresponds to eQTLs colocalizing with a GWAS but with a
644 different effect sign. This distinction enables the inference of the beta coefficient of the GWAS variant
645 based on the color of the eQTL beta.

646 For example, in B-cells (a), a decrease in the expression of the SLC22A5 gene is correlated with a
647 decrease in Asthma predisposition and an increase in extreme height and inflammatory bowel
648 diseases.

649

650 **Supplementary Figure 2: Comparison with the analysis by Watanabe et al 2019.**

651 a) Comparative analysis of trait categories counts in our study and that conducted by Watanabe et al.
652 [5].

653 b) The percentage of our eQTLs in the study by Watanabe et al. is depicted for various pleiotropy
654 levels [5]. The boxplots represent quartiles of the dataset, with whiskers extending to show the
655 remainder of the distribution except for outliers.

656 In figure (a), we conducted the Mann-Whitney U test. Significance levels are denoted as follows: ns
657 (not significant) for ≥ 0.05 , * for < 0.05 , ** for < 0.01 , *** for < 0.001 , and **** for < 0.0001 .

658

659 **Supplementary Figure 3: Pleiotropic eQTL distribution in allergic and blood-related and**
660 **cardiovascular diseases.**

661 These bar plots depict the proportions of eQTLs with different levels of pleiotropy that colocalize
662 with allergic (a) and cardiovascular diseases (b). The values above the bar plots show the count of
663 eQTLs.

664

665 **Supplementary Figure 4: eQTL and GWAS effects for given variant frequencies and sample**
666 **sizes.**

667 Average of the absolute effect of variants on gene expression (a, c) and traits (b, d) for variants with
668 frequencies in the interval between 0.45 and 0.55 (a, b) and samples size of the GWAS studies in the
669 interval between 75,000 and 125,000 (c, d).

670

671 **Supplementary Figure 5: Distribution of tissue and gene count.**

672 (a, c) Cumulative proportion of tissue count per eQTL-gene pair with increasing counts of tissues for
673 unconstrained variant frequency (a) and variant frequencies between 0.45 and 0.55. (b) Cumulative
674 proportion of gene count per eQTL-tissue pair with increasing counts of genes for variant frequencies
675 between 0.45 and 0.55.

676

677 **Supplementary Figure 6: Enrichment of pleiotropic eQTLs within CTCF ChIA-PET loops.**

678 These bar plots depict the odds ratio of pleiotropic eQTLs within CTCF ChIA-PET loops in isogenic
679 replicates r1 (a, c, e, g, j, l), r2 (b, d, f, h, k, m) and r3 (i) of cell lines A549 (a, b), CD8 T-cells (c, d),
680 GM10248 (e, f), HUVEC (g-i), K562 (j, k) and WTC11 (l-m).

681

682 **Supplementary Table 1: List of eQTL studies and annotations.**

683 This tables shows additional annotation regarding tissue category terms to complement the
684 information derived from eQTL studies in the EBI eQTL Catalogue.

685

686 **Supplementary Table 2: List of GWAS studies and annotations.**

687 The 417 GWAS studies in our analysis have been annotated with both a GWAS trait ontology and
688 GWAS trait category information. This additional annotation provides a more comprehensive
689 understanding of the traits associated with the genetic variations studied in the GWAS datasets.

690

691 **Supplementary Table 3: Proportion of leading SNPs in GWAS studies explained by colocalized
692 eQTLs.**

693 This table provides the percentage of leading SNPs in GWAS that colocalize with eQTLs.

694

695 **Supplementary Table 4: List of 5,345 GWAS variants that colocalize with eQTLs at
696 PP.H4.abf \geq 0.75 and SNP.H4.PP \geq 0.5 with annotations.**

697 This table offers the list of 5,345 GWAS variants that colocalize with eQTLs at PP.H4.abf \geq 0.75 and
698 SNP.H4.PP \geq 0.5 with annotations, including details such as the most cited eQTL gene and its number
699 of Pubmed citations (Columns H, I), trait category and trait counts (Columns J, K), a list of trait
700 categories, trait ontology, and eQTL gene identifier and symbols (Columns L, M), eQTL gene
701 identifier and symbols, and the number of eQTL genes (Columns N-P), the biological sample
702 category list and their numbers (Columns Q, R), the most cited gene identifier (Column S), and the
703 number of domains (equivalent to trait categories) in [5] (Column T).

704

705 **Supplementary Table 5: List of genomic regions containing colocalized eQTLs and GWAS
706 variants.**

707 This table contains the full list of regions, including information on the number and list of trait
708 category counts (Columns E, F), the symbol and identifier of eQTL genes (Columns G, I), and the
709 category of the tissue (Column H). The number of trait categories is determined by the count of
710 different trait categories in the region[18].

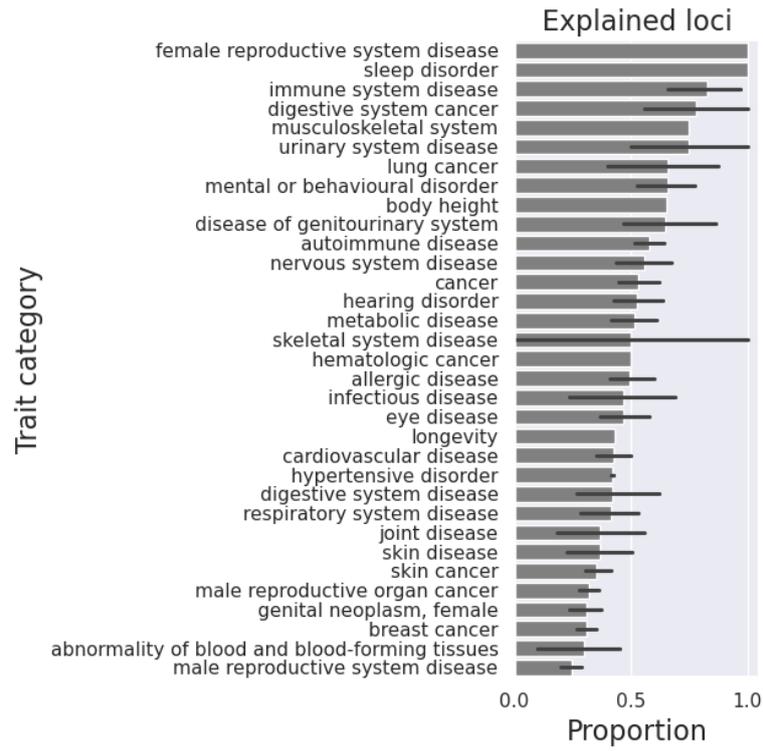
Chrom.	Pos (hg38)	Cytoband	rsid	Gene marker	Trait categories
1	8,437,247	1p36.23	rs301802	ENO1	allergic disease; cardiovascular disease; hypertensive disorder; mental or behavioural disorder
4	102,267,552	4q24	rs13107325	SLC39A8	autoimmune disease; cardiovascular disease; hypertensive disorder; musculoskeletal system
5	132,461,886	5q31.1	rs2522051	IL4	breast cancer; cardiovascular disease; hypertensive disorder; respiratory system disease
6	90,267,049	6q15	rs72928038	BACH2	allergic disease; autoimmune disease; metabolic disease; skin cancer
12	111,621,753	12q24.12	rs11065979	ALDH2	allergic disease; cardiovascular disease; eye disease; longevity
15	74,770,056	15q24.1	rs11072508	CYP1A1	cardiovascular disease; hypertensive disorder; immune system disease; joint disease
17	39,895,095	17q21.1	rs8067378	ORMDL3	allergic disease; autoimmune disease; cardiovascular disease; metabolic disease
19	48,703,417	19q13.33	rs601338	FUT2	autoimmune disease; cardiovascular disease; digestive system disease; hypertensive disorder

Table 1

Chr.	Start	End	Cytob.	Marker eQTL gene	Trait categories	Length	eQTLs cnt.
5	132,239,645	132,497,907	5q31.1	IL4	allergic disease; autoimmune disease; body height; breast cancer; cancer; cardiovascular disease; hypertensive disorder; respiratory system disease; skin disease	258,263	37
9	21,950,524	22,207,038	9p21.3	CDKN2A	breast cancer; cancer; cardiovascular disease; eye disease; genital neoplasm, female; male reproductive organ cancer; skin cancer	256,515	9
9	133,242,881	133,278,537	9q34.2	ABO	abnormality of blood and blood-forming tissues; cardiovascular disease; eye disease; genital neoplasm, female; infectious disease; mental or behavioural disorder	35,657	9
11	65,488,118	65,638,129	11q13.1	NEAT1	cardiovascular disease; eye disease; hypertensive disorder; joint disease; metabolic disease; respiratory system disease	150,012	8
11	65,747,403	65,909,045	11q13.1	RELA	allergic disease; autoimmune disease; breast cancer; cardiovascular disease; hypertensive disorder; skin cancer; skin disease	161,643	34
15	74,751,897	74,843,920	15q24.1	CYP1A1	breast cancer; cardiovascular disease; hearing disorder; hypertensive disorder; immune system disease; joint disease	92,024	15
15	78,428,581	78,762,558	15q25.1	CHRNA5	breast cancer; cardiovascular disease; longevity; lung cancer; mental or behavioural disorder; respiratory system disease	333,978	29
15	90,861,475	91,070,064	15q26.1	FURIN	breast cancer; cardiovascular disease; genital neoplasm, female; hypertensive disorder; mental or behavioural disorder; metabolic disease	208,590	38
16	89,626,691	89,808,935	16q24.3	MC1R	breast cancer; cancer; cardiovascular disease; eye disease; hypertensive disorder; skin cancer	182,245	24
17	2,006,529	2,311,037	17p13.3	SMG6	allergic disease; cardiovascular disease; eye disease; hypertensive disorder; mental or behavioural disorder; respiratory system disease	304,509	22
17	7,452,302	7,718,459	17p13.1	TP53	cancer; cardiovascular disease; hypertensive disorder; joint disease; respiratory system disease; skin cancer	266,158	27
20	63,588,387	63,857,282	20q13.33	EEF1A2	allergic disease; autoimmune disease; cancer; cardiovascular disease; hypertensive disorder; male reproductive organ cancer; metabolic disease; skin disease	268,896	30

Table 2

a



b

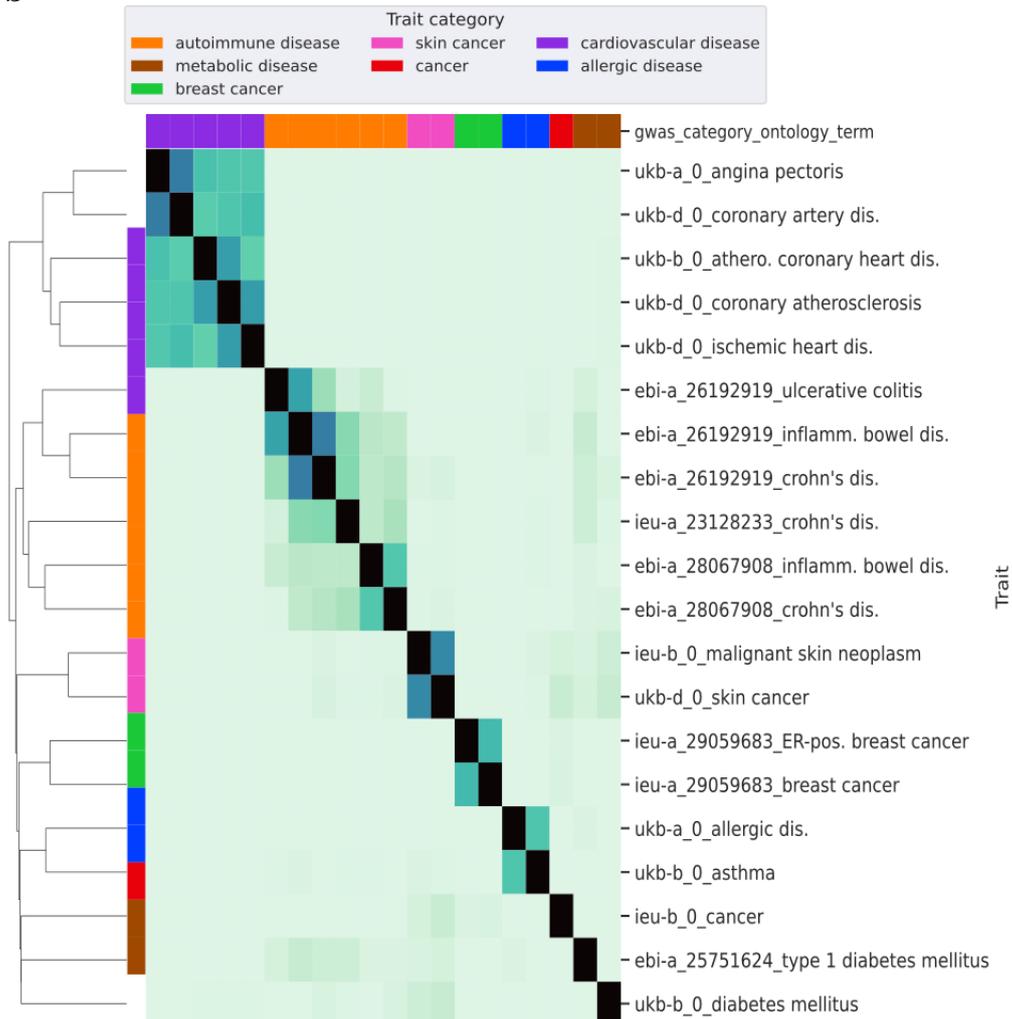


Figure 1

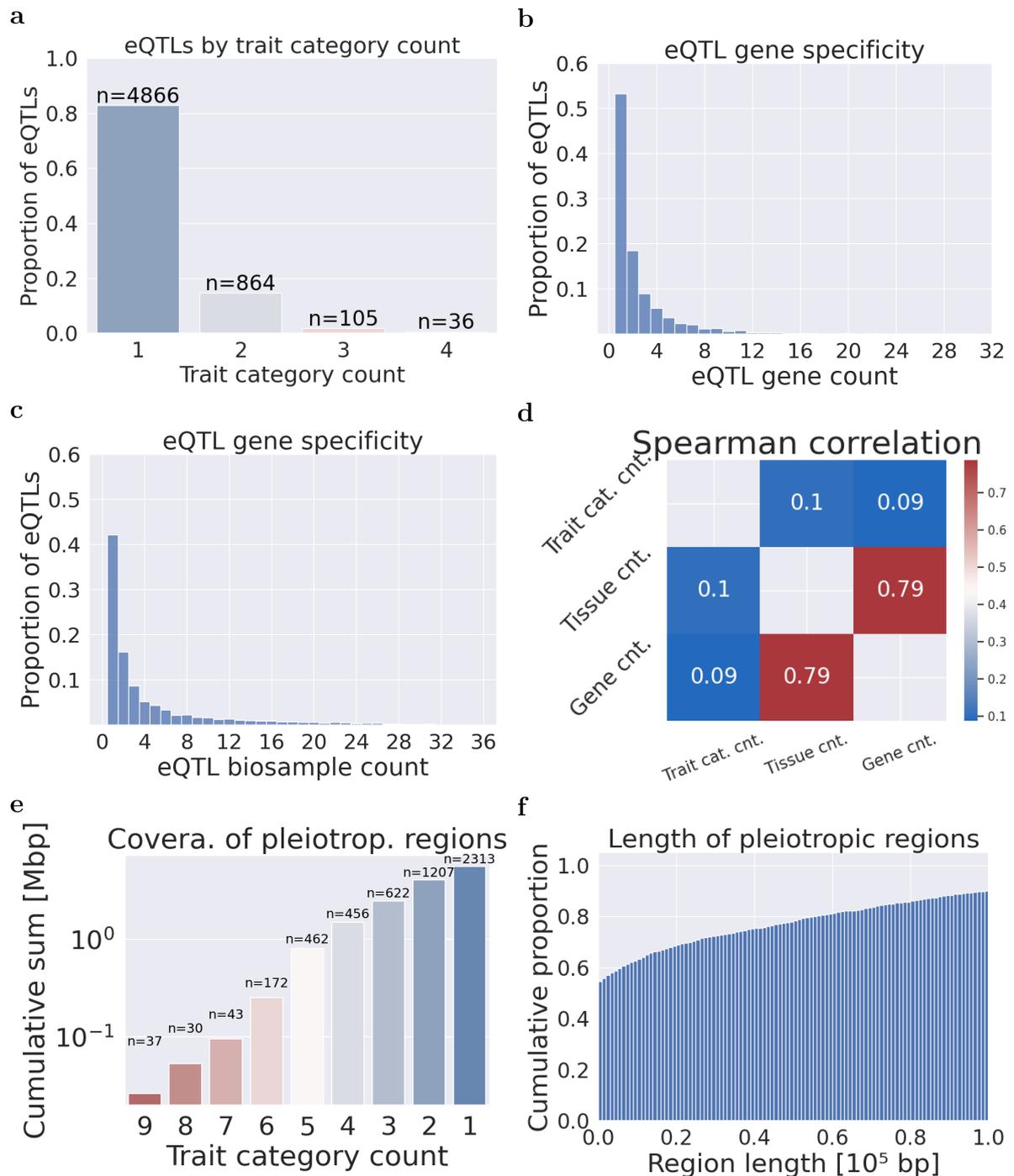


Figure 2

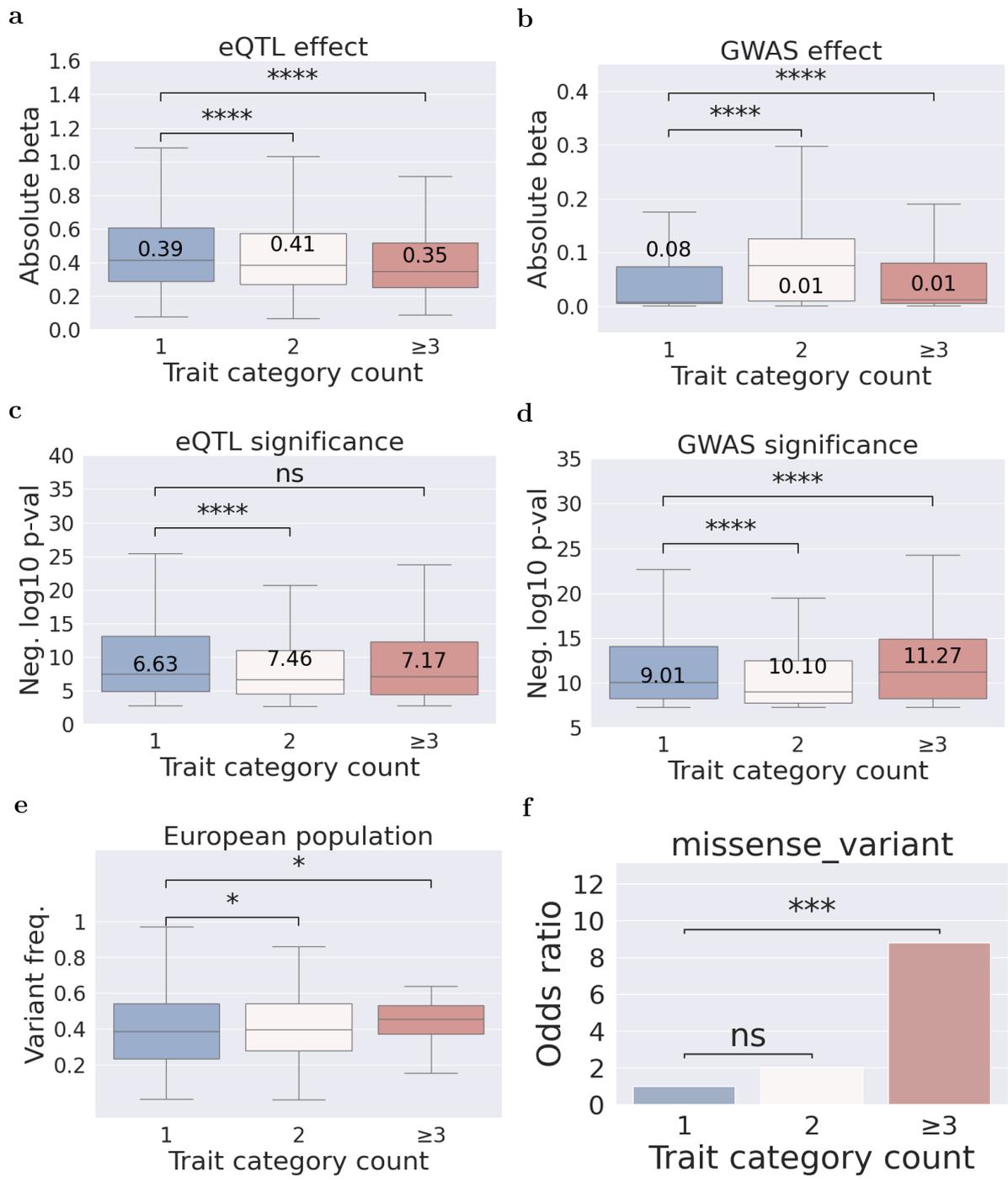


Figure 3

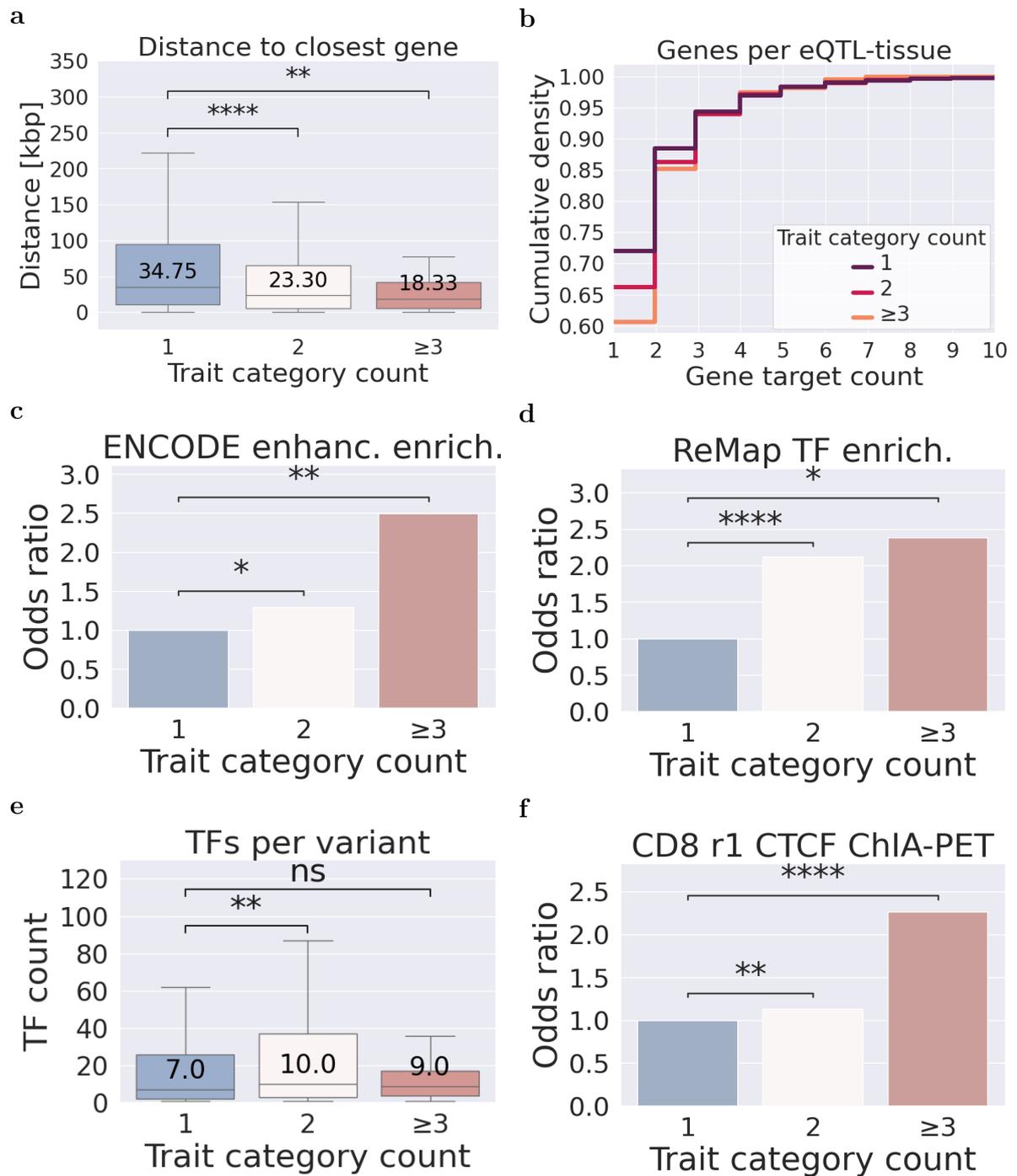
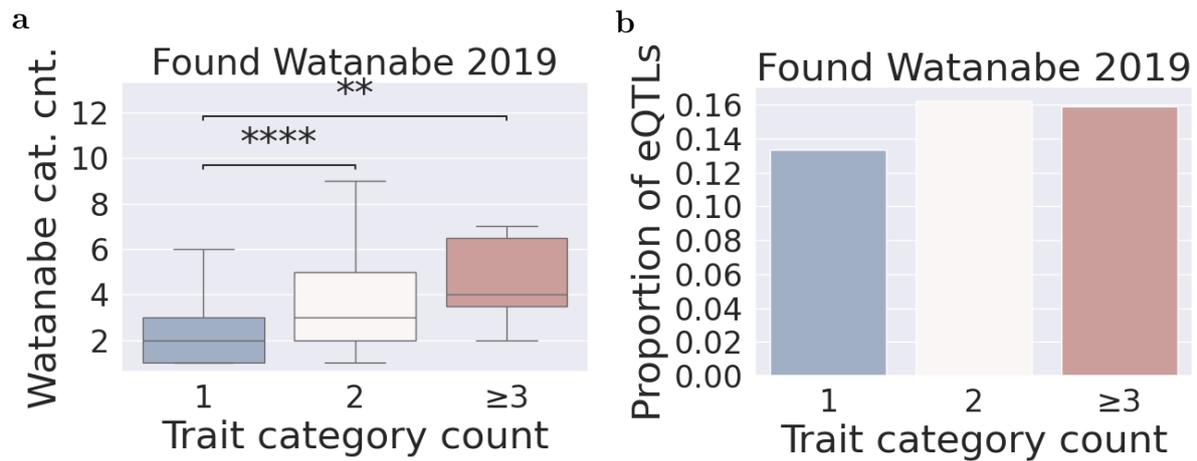
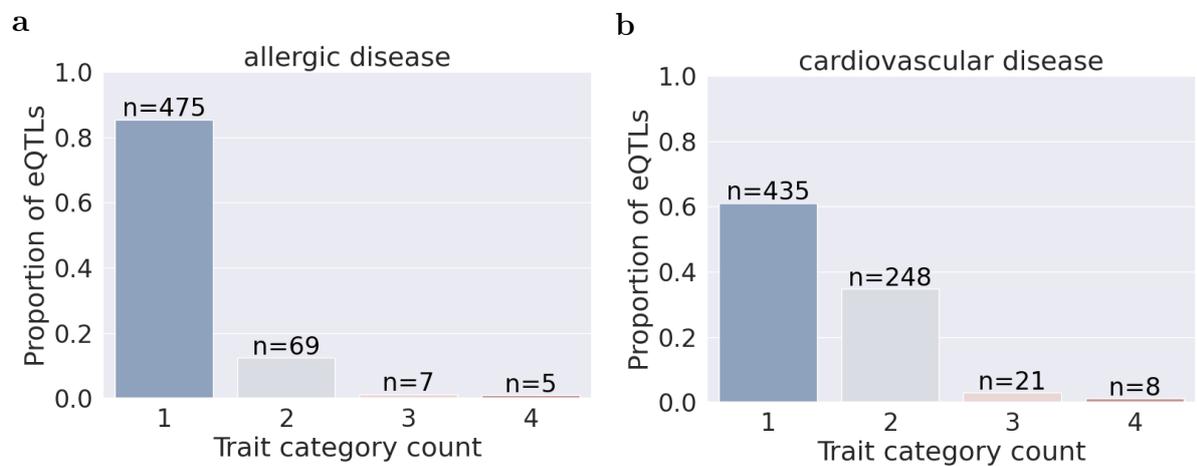


Figure 4

Supplementary Information

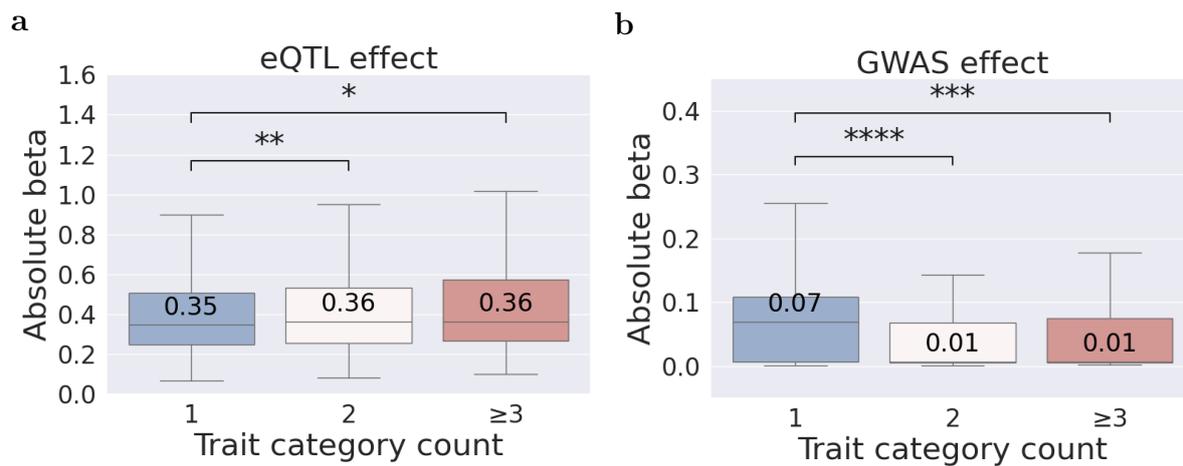


Supplementary Figure 2

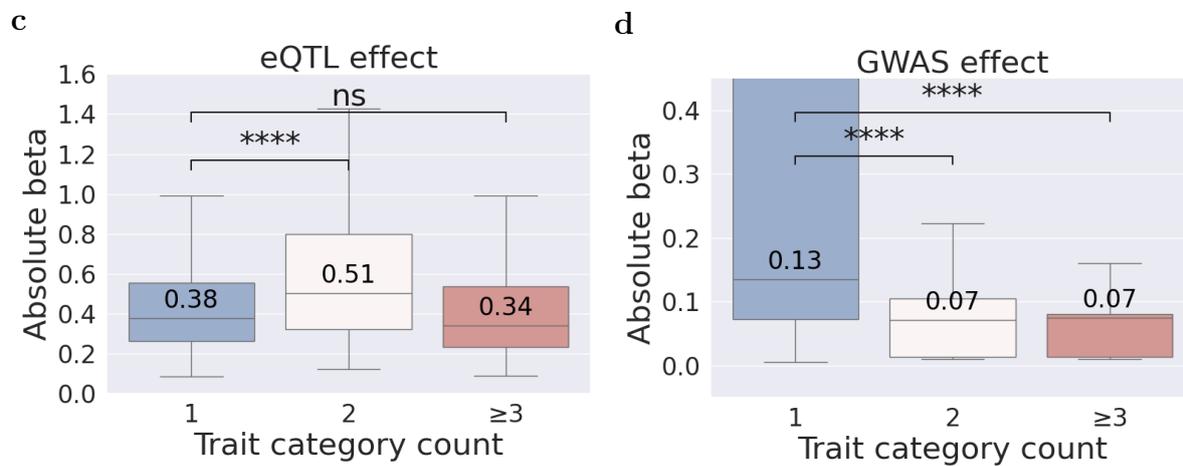


Supplementary Figure 3

Variant frequency: 0.45-0.55

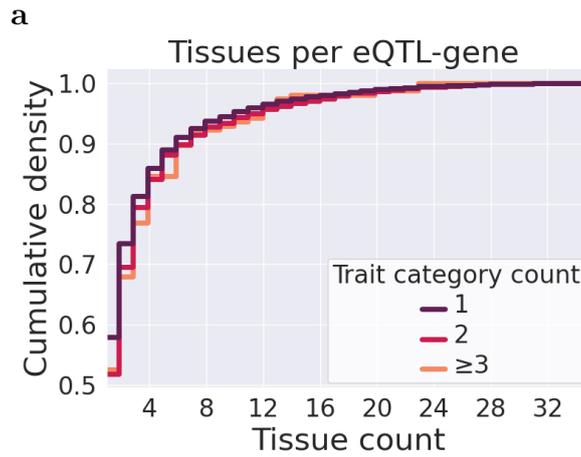


Sample size: 75,000-125,000

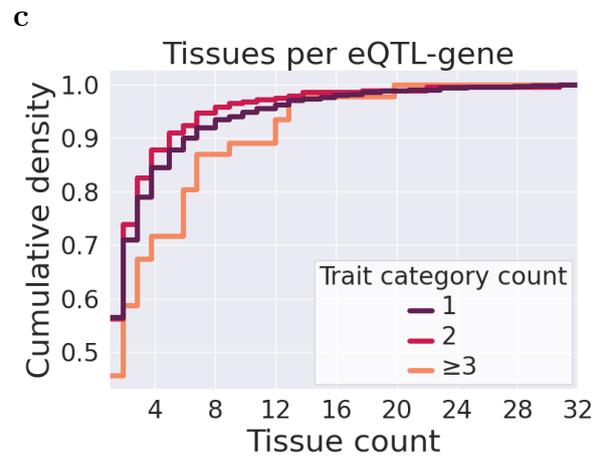
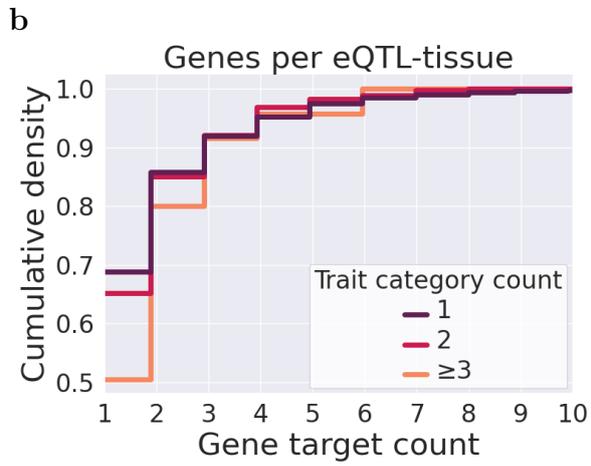


Supplementary Figure 4

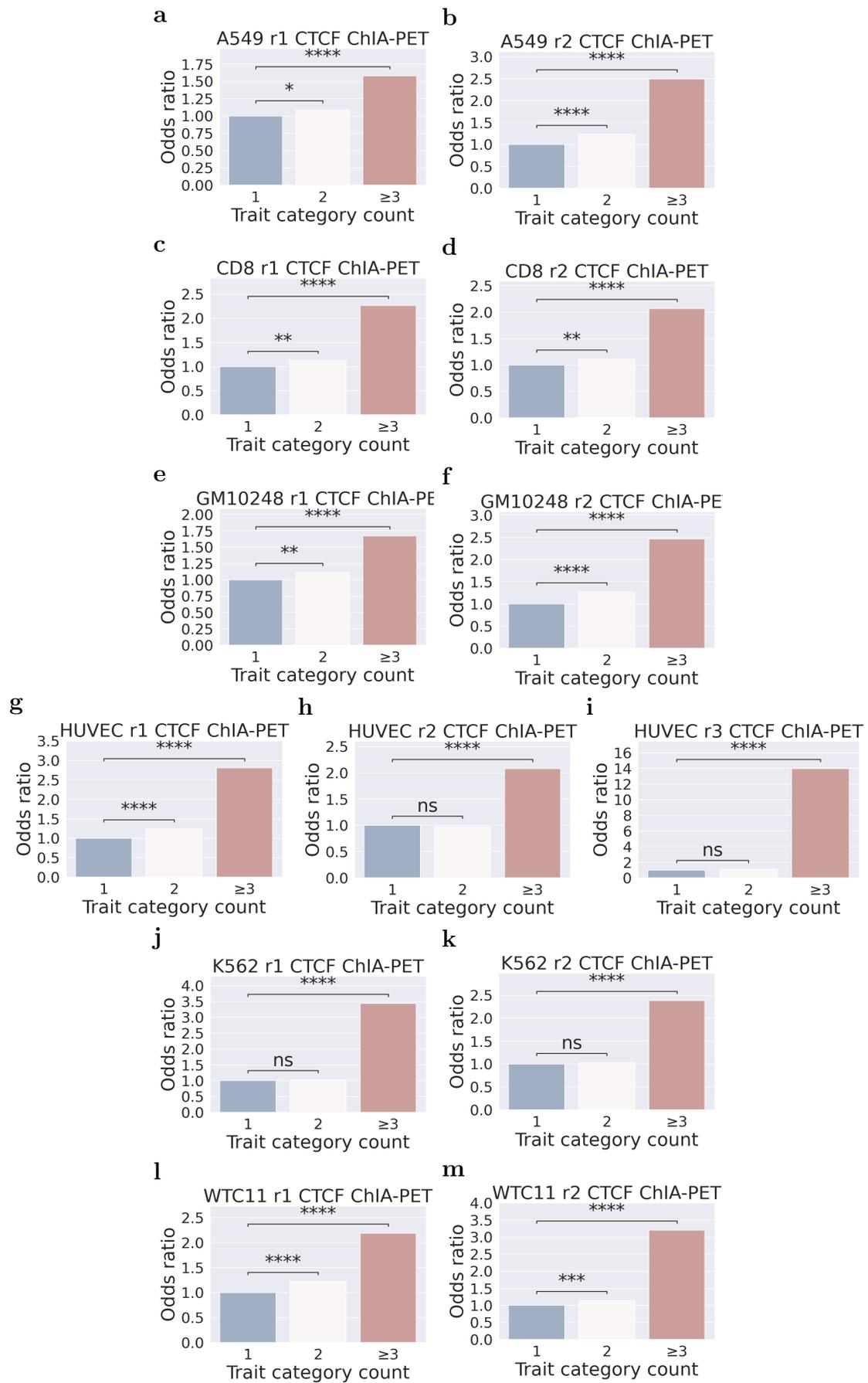
Variant frequency: 0-1



Variant frequency: 0.45-0.55



Supplementary Figure 5



Supplementary Figure 6