nature genetics

Article

https://doi.org/10.1038/s41588-026-02564-4

Multi-ancestry genome-wide association study of severe pregnancy nausea and vomiting

Received: 20 November 2024

Accepted: 26 February 2026

Published online: 14 April 2026

Check for updates

Marlena Fejzo[1] ✉, Xinran Wang[1], Qing Tan[1], Julia Zöllner[2,3], Natàlia Pujol-Gualdo[4], Triin Laisk[4], Estonian Biobank Research Team*, Sarah Finer[3], David A. van Heel[5], Health Research Team*, Ben Brumpton[6,7,8], Laxmi Bhatta[6,9,10], Kristian Hveem[6,7,11], Elizabeth A. Jasper[12], Digna R. Velez Edwards[12], Jacklyn N. Hellwege[12], Todd Edwards[12], Gail P. Jarvik[13], Yuan Luo[14], Atlas Khan[15], Kimber MacGibbon[16], Yuan Gao[17], Gaoxiang Ge[17], Inna Averbukh[18], Erin Soon[18], Michael Angelo[18], Per Magnus[19], Stefan Johansson[20], Pål R. Njølstad[20], Artem Kim[1], Steven Gazal[1], Marc Vaudel[20,21] ✉, Chang April Shu[1] ✉ & Nicholas Mancuso[1] ✉

Most pregnancies are affected by nausea and vomiting, but the most severe form—hyperemesis gravidarum—can be life threatening. Here we performed a multi-ancestry genome-wide association study of hyperemesis gravidarum in 10,974 cases and 461,461 controls across European, Asian, African and Latino ancestries. We identified ten associations: four identified previously (*GDF15*, *IGFBP7*, *PGR* and *GFRAL*) and six additional loci (*SLITRK1*, *SYN3*, *IGSF11*, *FSHB*, *TCF7L2* and *CDH9*). Downstream analyses revealed *GDF15* and *TCF7L2* expression primarily in extravillous trophoblasts, with opposing effects for *GDF15* between maternal and fetal genotype. Conversely, *IGFBP7* and *PGR* were expressed primarily in maternal spiral arteries, with effects limited to the maternal genome. Selected loci were associated with abnormal pregnancy weight gain, duration, birth weight and pre-eclampsia. Functional studies identified additional associations including antisense *IGFBP7-AS1* and protein ACP1. Potential roles for candidate genes in appetite, insulin signaling and brain plasticity provide pathways to explore etiological mechanisms and therapeutic avenues.

Most pregnancies are affected by nausea and vomiting (NVP), but in 0.3–10.8% of pregnancies the symptoms can be severe enough to cause maternal weight loss and adverse maternal and fetal outcomes[1,2]. The most severe form, hyperemesis gravidarum (HG), is the leading cause of hospitalization in the first trimester and the second leading cause, after preterm labor, of pregnancy hospitalization overall[3]. Current treatments for HG are frequently ineffective in improving patient symptoms, thereby increasing risk of pregnancy termination, suicidal ideation, postpartum depression and numerous other maternal and offspring comorbidities[4–9]. Therefore, understanding of HG etiology is critical to begin to address the negative impact severe NVP has on maternal and child health.

Although historical hypotheses have centered around human chorionic gonadotropin (hCG), recent large-scale genetic studies have implicated the *GDF15* gene encoding growth differentiation factor-15—a hormone associated with nausea and vomiting[10,11]. *GDF15* was identified as the greatest genetic risk factor for HG in both a genome-wide and an exome-wide association study, and a rare mutation in *GDF15* was associated with a greater than tenfold increased risk for HG[10,11]. Other significant genetic associations that have been replicated include the gene coding for the brainstem-restricted receptor for GDF15, GDNF family receptor alpha-like (*GFRAL*) and the placental proteins insulin-like growth factor binding protein 7 (*IGFBP7*) and progesterone receptor (*PGR*)[10,12]. Despite these substantial advances, further work in larger

A full list of affiliations appears at the end of the paper. ✉e-mail: marlena.fejzo@med.usc.edu; marc.vaudel@uib.no; april.shu@med.usc.edu; nicholas.mancuso@med.usc.edu

https://doi.org/10.1038/s41588-026-02564-4

Table 1 | Source populations and ancestries

Source	Case (N1)	Control (N2)	EUR (N1/N2)	AFR (N1/N2)	Asian (N1/N2)	AMR (N1/N2)	UNK[a] (N1/N2)
23andMe, Inc.	1,306	15,756	1,306/15,756	0/0	0/0	0/0	0/0
HER	926	660	804/615	32/11	14/11	58/17	18/6
FinnGen R8	2,092	163,702	2,092/163,702	0/0	0/0	0/0	0/0
EstBB	3,536	32,113	3,536/32,113	0/0	0/0	0/0	0/0
UKBB	220	180,917	220/180,917	0/0	0/0	0/0	0/0
Genes and Health	1,544	11,128	0/0	0/0	1,544/11,128	0/0	0/0
HUNT	269	36,460	269/36,460	0/0	0/0	0/0	0/0
eMERGE	504	6,364	275/4,897	197/1,037	32/430	0/0	0/0
MoBa	577	14,361	577/14,361	0/0	0/0	0/0	0/0
Total	10,974	461,461	9,079/448,821	229/1,048	1,590/11,569	58/17	18/6

[a]Unknown (Other) ancestry reported using Regeneron Genetics Pipeline[11]. The nine source populations (23andMe, Inc., HER, FinnGen R8, EstBB, UKBB, Genes and Health, HUNT, eMERGE and MoBa) included in the study and the corresponding number of cases and controls in total and by ancestry are shown. Ancestries include European (EUR), African (AFR), Asian (East and South Asian combined), Admixed American (AMR), unknown (UNK). N1, number of cases; N2, number of controls.

Table 2 | Lead SNPs for multi-ancestry meta-analysis of nine GWASs and European-only meta-analysis of six GWASs

Index SNP	Chr.	GRCh38 position	Closest gene	A1/A0	Beta	s.e.	Multi-ethnic *P*	Direction	Heterogeneity *P*	Known/X	European-only *P*
rs1058587	19	18,388,612	GDF15	C/G	0.2783	0.0181	3.84×10^{-53}	+++++++?+	0.0042	Known	4.13×10^{-40}
rs9312688	4	57,486,707	IGFBP7	A/G	0.1569	0.0162	4.09×10^{-22}	++?++++?+	0.009924	Known	5.65×10^{-20}
rs12790159	11	101,338,299	PGR	T/C	0.2072	0.0231	3.47×10^{-19}	++??+++?+	0.04704	Known	1.60×10^{-12}
rs10948901	6	55,324,684	GFRAL	T/C	0.0972	0.0157	5.59×10^{-10}	++?+−++?+	0.04469	Known	5.81×10^{-10}
rs7101406	11	30,144,159	FSHB	T/C	−0.0946	0.0157	1.80×10^{-9}	−?−−−−−	0.5996	X	2.88×10^{-7}
rs17077610	13	83,977,056	SLITRK1	A/C	−0.1547	0.027	1.04×10^{-8}	−?−+−+?−	0.2689	X	9.14×10^{-7}
rs5994661	22	33,050,220	SYN3	A/G	0.0875	0.0153	1.19×10^{-8}	++?+−++++	0.06154	X	5.69×10^{-7}
rs56337209	3	118,525,324	IGSF11	A/G	0.0888	0.0157	1.69×10^{-8}	++?++++?+	0.9245	X	3.49×10^{-7}
rs76856932	10	113,347,748	HABP2[a]	A/C	0.3236	0.0581	2.48×10^{-8}	++?+−+-??	0.03454	X	1.63×10^{-7}
rs10073299	5	27,647,394	CDH9	T/C	2.1933	0.4006	4.37×10^{-8}	?+????+??	0.5655	X	0.1282

[a]rs76856932 maps between *HABP2* and *TCF7L2*. Expression of *TCF7L2* in placenta and thalamus and published roles in glucose sensing and appetite indicate that it is a more likely candidate gene than nearest gene *HABP2*. This table shows the lead SNP with most significant association (Index SNP), corresponding chromosome (Chr.), position of lead SNP using reference RefSeq assembly hg38 (GRCh38), gene with transcription start site closest to index SNP, effect allele (A1), alternate allele (A0), effect size (Beta), s.e., probability value (*P*) for the whole multi-ethnic cohort and separately the European-only cohort, with direction of effect positive (+), negative (−) or unknown (?), listed from left to right corresponding to cohorts Finngen, 23andMe, HER, EstBB, UKBB, Genes and Health, HUNT, eMERGE and MoBa. There is no heterogeneity between studies (Heterogeneity *P*). Known/X indicates whether the locus has been published previously (Known) or identified in this study (X). MAFs are shown in Supplementary Table 2.

and more diverse populations is necessary to determine the generalizability of these findings and identify additional genetic associations and potential therapeutic targets.

Here we performed a multi-ancestry genome-wide association study (GWAS) of 10,974 HG/excessive vomiting in pregnancy cases and 461,461 controls across European, Asian, African, and Latino ancestries from nine contributing studies. GWAS confirmed previous GWASs with *GDF15*, *GFRAL*, *IGFBP7* and *PGR*, and identified six additional loci (SLIT- and NTRK-like family member 1 (*SLITRK1*), synapsin-3 (*SYN3*), immunoglobulin superfamily member 11 (*IGSF11*), follicle stimulating hormone (FSH) subunit beta (*FSHB*), transcription factor 7-like 2 (*TCF7L2*) and cadherin-9 (*CDH9*)) associated significantly with HG risk. We applied functional studies and analyzed the risk loci with respect to maternal and fetal/paternal contributions, other phenotypic associations, anthropometric measures and pregnancy outcomes, temporal association with NVP throughout gestation and spatiotemporal expression in the developing placenta. Overall, our work provides additional insight into the genetic etiology of HG risk.

Results

HG GWAS meta-analysis identifies additional risk regions

Our discovery meta-analysis of GWAS on HG encompassed nine independent studies: three cohorts from the United States (23andMe, Inc.[10], Hyperemesis Education and Research (HER)[11] and Electronic Medical Record and Genomics (eMERGE)[13]), one from Estonia (Estonian Biobank (EstBB))[14], two from the United Kingdom (UK Biobank (UKBB)[15] and Genes and Health[16]), one from Finland (FinnGen Biobank, release v.8 (ref. 17)) and two from Norway (HUNT[18] and The Norwegian Mother, Father and Child Cohort Study (MoBa)[19]). In total, we analyzed 10,974 cases and 461,461 controls across European ($N = 457{,}900$), Asian ($N = 13{,}159$), African ($N = 1{,}277$) and Latino ($N = 75$) ancestries (Table 1). Broadly, cases were defined as patients diagnosed with HG or excessive vomiting during pregnancy, and controls were defined as women with no NVP diagnoses (Supplementary Table 1). For each contributing study we performed rigorous quality control (QC) on genotyped and imputed variants, summary statistics and outlier substudies (that is, studies contributing to individual biobank results; Methods). To account for population structure and reflect the predominance of European participants, we conducted an additional meta-analysis restricted to European cohorts (FinnGen, 23andMe, EstBB, UKBB, HUNT and MoBa; Table 2).

We first performed genome-wide association scans for the seven contributing studies with individual-level data (Supplementary Table 1) using logistic mixed-effect models. Next we performed a fixed-effects meta-analysis using summary statistics from autosomes of all nine cohorts and sex chromosomes where available, resulting in a total

https://doi.org/10.1038/s41588-026-02564-4

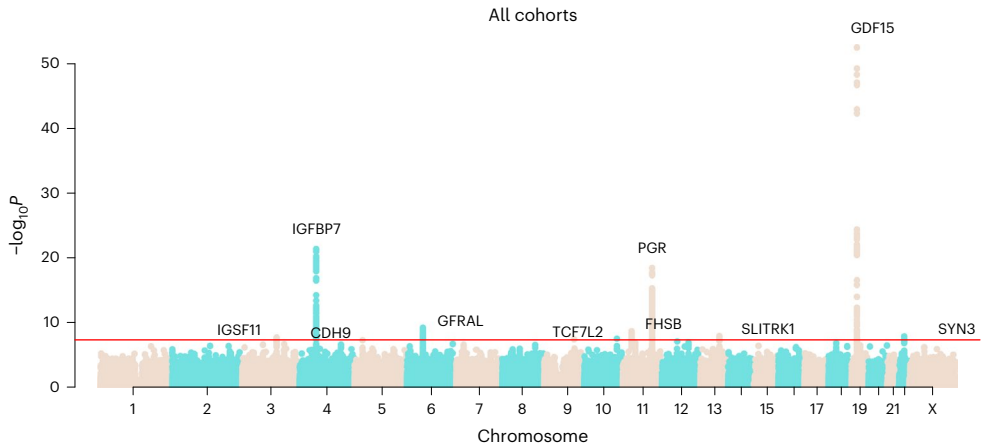


Fig. 1 | Multi-ancestry genome-wide association scan for HG. Manhattan plot showing distribution of association test statistics ($-\log_{10}$ transformed P value against genomic positions) for the binary phenotype HG/excessive vomiting cases versus unaffected controls (no HG)[79]. Chromosomes are arranged along the x axis. $\log_{10}$-scaled P values are shown on the y axis. The red line represents genome-wide significant threshold $P < 5 \times 10^{-8}$. The significant SNPs are annotated with the nearest candidate gene.

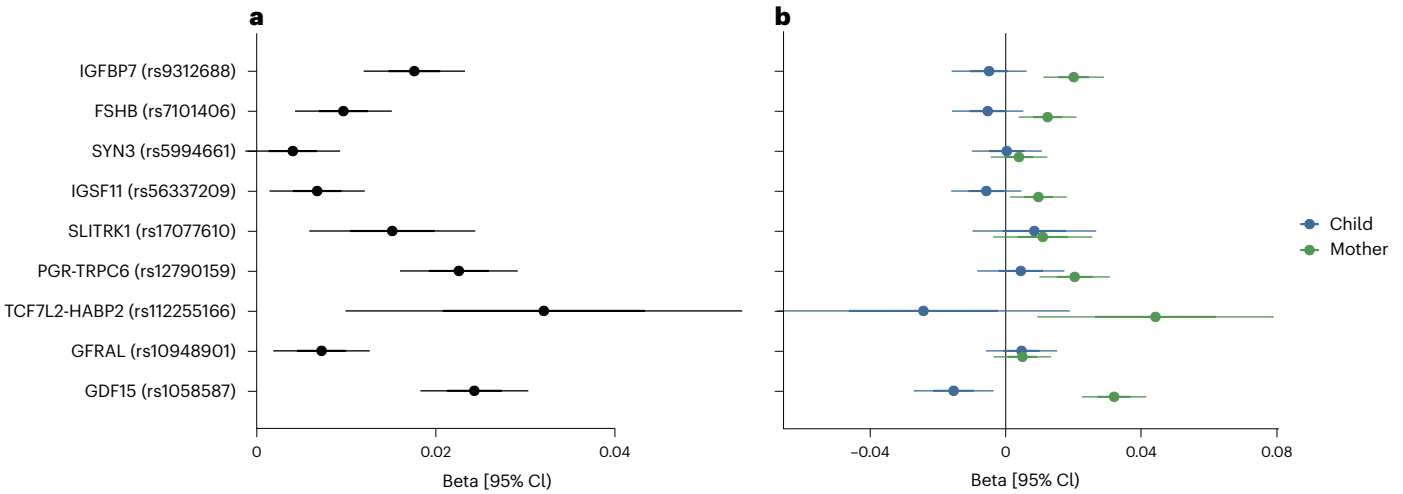


Fig. 2 | Effect size estimates for the lead SNP at each locus in the MoBa cohort. **a**,**b**, Forest plot of the effect size estimates when encoding no NVP, NVP and HG as 0, 1 and 2, respectively, for maternal effect estimates as obtained from the GWAS against the maternal genome (**a**) and conditioned maternal and fetal effects obtained after weighted linear model adjustment of effect estimates in GWASs against the fetal, maternal and paternal genomes (**b**). For the sake of readability, paternal effect sizes are not represented. Points and error bars in **a** represent the effect size estimates and 95% CIs of the association with the maternal genome; $N = 56{,}225$ mothers. Points and errors bars in **b** represent effect size estimates and 95% CIs of the conditioned maternal and fetal effects; $N = 56{,}225$ mothers and $N = 73{,}389$ children.

of 43,961,852 examined variants (Methods). Overall, we identified ten independent genome-wide significant loci ($P < 5 \times 10^{-8}$; Fig. 1). Of these ten loci, six were not identified previously (defined as at least 1 Mb distance from any other genome-wide significant association; Fig. 1, Table 2, Supplementary Table 2 and Supplementary Figs. 1–10). No variant demonstrated genome-wide significance for heterogeneity across studies ($P < 5 \times 10^{-8}$; Table 2; Methods). Then, we performed European-only meta-analysis, restricted to six European cohorts, with a total of 38,769,878 variants. We identified six genome-wide significant signals ($P < 5 \times 10^{-8}$). We were able to replicate nine out of ten significant hits from multi-ancestry meta-analysis, except for rs10073299 on chromosome (Chr.) 5, at a relaxed threshold of $P < 5 \times 10^{-6}$. Among these, four hits were significant at genome-wide significance levels of $P < 5 \times 10^{-8}$ (Table 2 and Extended Data Figs. 1 and 2). Next, we performed statistical fine mapping for each significant 1-Mb region to identify independent signals and putative causal variants. Briefly, we used susieR to perform statistical fine mapping of meta-analysis summary statistics at identified risk regions, using estimates of linkage disequilibrium (LD) from 1000 Genomes Phase3 unrelated Europeans[20,21]. Before fine mapping, we pruned summary statistics to sites having minor allele frequency (MAF) ≥ 0.001 and nonambiguous alleles (for example, A/T, C/G) with exceptions for lead variants (Methods). Fine mapping was not performed at the risk region around *CDH9* (rs10073299, chr. 5: 27647394) due to GWAS signals being pruned at MAF thresholds > 0.001. Across nine out of ten risk regions, fine mapping identified 14 95% credible sets, with seven of nine fine-mapped regions containing a single credible set, suggesting mild allelic heterogeneity underlying HG risk regions (Supplementary Tables 2–4 and Extended Data Figs. 3 and 4). Credible sets contained 20 single nucleotide polymorphisms (SNPs) on average, consistent with pervasive LD patterns at GWAS risk regions; however, 3 of 14 credible sets were composed of an individual SNP. In addition, to access population-specific signals, we performed a separate fine mapping using meta-analysis summary statistics from the European-only meta-analysis. We identified 12 95% credible sets across nine out of ten risk regions and a single credible set was identified in seven of these regions. On average, there are 25 SNPs per credible

https://doi.org/10.1038/s41588-026-02564-4

Table 3 | Summary of FUSION TWAS results

Gene	Chr.	*N*	Minimal *P* value	Tissue with minimal *P* value	Tissue list
ACP1	2	2	5.59×10^{-6}	Brain_cerebellum	Brain, esophagus
SH3YL1	2	7	3.05×10^{-6}	Heart_left_ventricle	Heart, esophagus, salivary gland, vagina, stomach
ALKAL2	2	1	1.63×10^{-6}	Brain_spinal_cord_cervical_c-1	Spinal cord
ENSG00000243276	3	1	1.36×10^{-5}	Cells_EBV-transformed_lymphocytes	EBV-transformed lymphocytes
IGFBP7-AS1	4	1	1.50×10^{-6}	Breast_mammary_tissue	Breast
ENSG00000251459	4	1	1.27×10^{-10}	Heart_left_ventricle	Heart
HERC3	4	1	4.80×10^{-6}	Brain_nucleus_accumbens_basal_ganglia	Brain
CAMLG	5	13	3.59×10^{-6}	Brain_anterior_cingulate_cortex_BA24	Brain, esophagus, breast, adipose, heart, spinal cord
ENSG00000271860	6	1	4.36×10^{-6}	Brain	Brain
LINC01301	8	4	4.63×10^{-6}	Brain_hypothalamus	Brain
LINC01616	11	1	2.88×10^{-6}	Pituitary	Pituitary
ARL14EP-DT	11	1	2.32×10^{-8}	Adipose_visceral_omentum	Adipose
ARL14EP	11	3	6.20×10^{-8}	Minor_salivary_gland	Salivary gland, stomach, breast
RPL12P30	11	2	1.12×10^{-8}	Minor_salivary_gland	Salivary gland, brain
ABCB9	12	1	7.67×10^{-6}	Brain_caudate_basal_ganglia	Brain
MTRFR	12	2	2.49×10^{-6}	Esophagus_muscularis	Esophagus, vagina
ENSG00000235423	12	5	1.64×10^{-6}	Breast_mammary_tissue	Breast, brain, spinal cord
CDK2AP1	12	6	8.71×10^{-7}	Esophagus_gastroesophageal_junction	Esophagus, stomach, adipose, heart, vagina, salivary gland
KMT5A	12	2	3.91×10^{-6}	Esophagus_mucosa	Esophagus, heart
ENSG00000273654	19	1	1.76×10^{-7}	Adipose_visceral_omentum	Adipose
PGPEP1	19	4	4.25×10^{-11}	Pituitary	Pituitary, stomach, whole blood, heart
POM121L9P	22	1	1.63×10^{-5}	Vagina	Vagina
UPB1	22	4	1.99×10^{-6}	Whole_blood	Whole blood, brain

FUSION TWAS was performed across 27 tissues from GTEx v.8 and 1 tissue from Gandal et al.[26]. The gene column lists symbols for genes that passed tissue-specific significance thresholds (0.05/number of genes tested in each tissue). Chr. for corresponding genes are listed. Number tissue indicates the count (*N*) of tissues in which each gene was significant and the tissue list column provides the complete set of tissues in which the gene met the significance threshold. For each gene, the minimal *P* value and the corresponding tissue with minimal *P* value are reported.

set and 2 of 12 credible sets composed of a single SNP. These findings are consistent with those from the multi-ancestry meta-analysis (Supplementary Tables 3–5 and Extended Data Figs. 3 and 4).

Opposing maternal and fetal contributions for GDF15

Next we investigated the relative contribution of maternal and fetal genetics to the etiology of HG in the MoBa study—a pregnancy-based prospective cohort representing more than 100,000 pregnancies. Briefly, we leveraged a recent approach that estimates the contribution of parental/fetal effects by analyzing GWASs conducted against parents and children separately (Methods)[22]. Among the ten genome-wide significant loci identified in the meta-analysis, nine were genotyped with high quality in MoBa and selected for downstream analysis.

For the SNPs with nearest genes *IGFBP7*, *FSHB*, *IGSF11*, *PGR* and *TCF7L2-HABP2*, the effect was carried primarily by the maternal genome. At our current sample size, the confidence interval (CI) of the fetal effect was crossing the null for all these loci (Fig. 2). We found opposite allele effects for rs1058587 in *GDF15* on HG risk between mother and fetus. This is in line with the hypothesis that low levels of circulating GDF15 before pregnancy expressed from a maternal GDF15-decreasing genotype and higher levels of GDF15 during pregnancy expressed from a fetal GDF15-increasing genotype can both contribute to HG risk due to a preconditioned maternal hypersensitivity to GDF15 and the rise of GDF15 in early pregnancy[23]. The maternal effect was stronger than the fetal effect, suggesting maternal prepregnancy conditioning to GDF15 may largely prevail over the fetal genetic contribution to GDF15 level during pregnancy.

HG risk is enriched for regions under constraint

To characterize the genetic architecture of HG, we performed partitioned LD score (LDSC) regression using the meta-analysis summary statistics (Methods). First, LDSC[24] estimated an intercept of 1.0281 (s.e.: 0.0086), indicating that results may be affected in part by population stratification. Second, LDSC estimated significant heritability on a liability scale ($h_g^2 = 0.12$, $P = 1.10\times10^{-17}$) when provided 2.3% HG case prevalence in our study samples and 2% population prevalence. To account for varying reports of population prevalence[1,25], we reanalyzed our results using differing values and found broadly consistent estimates of h_g^2 on the liability scale (Supplementary Table 6). When partitioning h_g^2 across various functional and evolutionary categories, we found significant associations between measures of gene constraint with HG heritability ($P < 0.0005$; Extended Data Fig. 5). For example, genomic evolutionary rate profiling[23] scores, suggesting highly disproportionate conservation and providing evidence that HG risk loci are under evolutionary constraint. A separate LDSC analysis restricted to European-only cohorts yielded similar results, with intercept of 1.0289 (s.e.: 0.0081), heritability remained significant on a liability scale ($h_g^2 = 0.11$, $P = 7.83\times10^{-16}$). The heritability and gene constraints remained significantly associated (Supplementary Table 5 and Extended Data Fig. 5).

Integrative analyses identify 23 genes linked with HG risk

To identify expression-trait associations, we performed a transcriptome-wide association study (TWAS) for precomputed expression weights on 27 tissues (adipose, brain/nervous system, cardiovascular,

https://doi.org/10.1038/s41588-026-02564-4

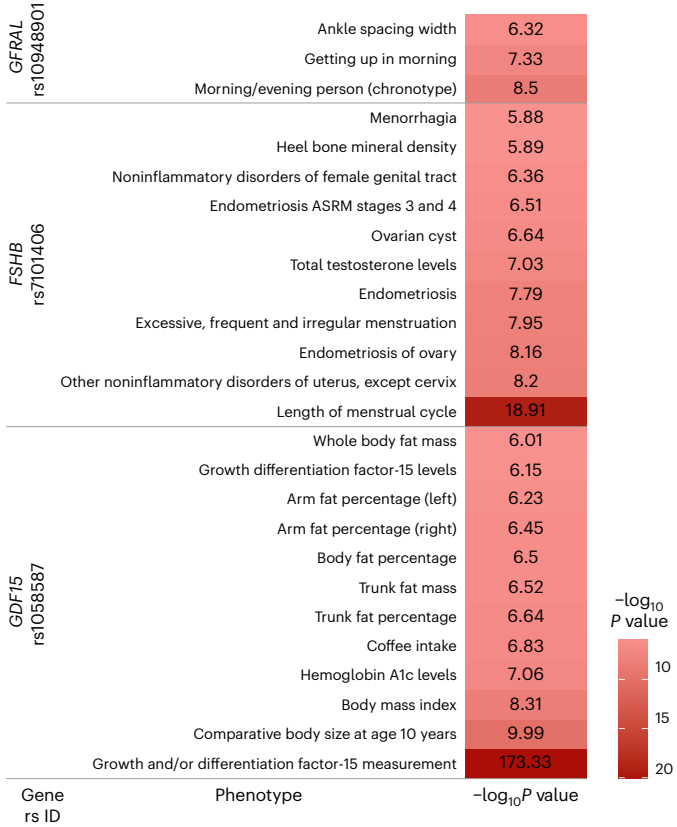


Fig. 3 | PheWAS of ten genome-wide significant SNPs in HG among 3,380 phenotypes in Open Target Genetics[80]. Significant phenotype associations were identified for three variants (rs1058587, rs10948901 and rs710140) after adjusting for multiple testing ($P < 1.48 \times 10^{-6}$). Darker red indicates a larger $-\log_{10}P$ value. ASRM, American Society for Reproductive Medicine.

reproductive, digestive, blood/immune and endocrine tissue) from Genotype-Tissue Expression (GTEX) Project v.8 and 1 brain tissue from Gandal et al. using meta-analysis summary statistics[26,27]. Among 149,484 gene–tissue tests across 28 tissues (22,744 distinct genes), 65 passed the tissue-specific significance threshold, representing 23 distinct genes. Of the 65 associations, 7 located within 250 kb of the ten genome-wide significant variants, representing four genes/RNA genes (*PGPEP1*, *ENSG00000251459*, *ARL14EP-DT* and *ENSG00000243276*) and 1 located near antisense RNA (*IGFBP7-AS1*). The greatest contributions are from pituitary (*PGPEP1*; $P = 4.25 \times 10^{-11}$), heart left ventricle (*ENSG00000251459*; $P = 1.27 \times 10^{-10}$), stomach (*PGPEP1*; $P = 2.02 \times 10^{-9}$), minor salivary gland (*PGPEP1*; $P = 1.12 \times 10^{-8}$) and adipose visceral omentum (*ARL14EP-DT*; $P = 2.32 \times 10^{-8}$; Table 3 and Extended Data Figs. 6 and 7).

We conducted a proteome-wide association study (PWAS) of HG using proteomics data from UKBB[28] and Xu et al.[29] and identified three proteins (GDF15, acid phosphatase 1 (ACP1) and leukocyte immunoglobulin-like receptor B2 (LILRB2)) associated significantly with HG after Bonferroni correction (Extended Data Fig. 8 and Supplementary Table 7). In the PWAS of HG using UKBB proteomics datasets, GDF15 reached proteomics-wide significance ($P = 2.48 \times 10^{-24}$), which is the strongest signal and consistent with our previous HG GWAS and exome-wide association study results[10,11]. In the PWAS of HG using proteomics datasets from Xu et al.[29], two proteins reached proteome-wide significance ($P < 3.07 \times 10^{-5}$): ACP1 ($P = 4.90 \times 10^{-6}$; Supplementary Table 7)—a low-molecular-weight phospho-tyrosine phosphatase involved in intracellular signaling—and LILRB2 ($P = 1.31 \times 10^{-5}$; Supplementary Table 7)—an inhibitory immune receptor expressed on myeloid cells.

We further explored the regulatory mechanisms of HG-associated variants by integrating single-cell assay for transposase-accessible chromatin (scATAC) using sequencing (scATAC–seq) peaks from placenta during early pregnancy and chromatin immunoprecipitation sequencing (ChIP–seq) peaks from adult female tissues (ENCODE)[30–33]. We identified scATAC accessibility signals within ±500 bp of rs7101406 (*FSHB*) and rs5994661 (SYN3), and ChIP–seq signals within a 10,000 bp region of rs1058587 (*GDF15*) in brain and chorionic villus and at rs5994661 (*SYN3*) in the brain (Supplementary Figs. 1–10).

Regulation predictions were identified for all risk loci using Haploreg v.4.2 (ref. 34) (Supplementary Table 8), with, of note, a DNase I-hypersensitive site in placenta for the *GDF15* locus (rs1058587) and an enhancer histone mark in placenta and brain for the *TCF7L2* locus (rs76856932).

HG risk genes linked to characteristics and outcomes

To understand the broader phenotypic consequences of HG risk variants identified in this study, we performed a phenome-wide association study (PheWAS) for each of the ten genome-wide significant SNPs in our meta-analysis using the OpenTargets platform (Methods)[35]. After adjusting for multiple testing ($P < 1.48 \times 10^{-6}$), three of ten variants exhibited associations with additional phenotypes, including traits involved in GDF15 measurements (rs1058587), body size (for example, fat percentage; rs1058587), female reproductive system (for example, such as length of menstrual cycle, endometriosis and ovarian cysts; rs7101406) and chronotype (for example, getting up in the morning; rs10948901; Fig. 3).

Next we sought to explore the phenotypic consequences of these variants focusing on maternal characteristics (maternal weight/weight gain/body mass index (BMI) at beginning of pregnancy, at 15 weeks' gestation and at term) in HG cases and controls in the MoBa cohort (Supplementary Table 9). Overall, we identified significant associations between risk loci and maternal weight and BMI at the beginning of pregnancy (*GDF15*, *SLITRK1*). Risk loci were also associated with maternal weight gain at 15 weeks (associated most strongly with *GDF15* but also with *IGFBP7* and *PGR*) and at the end of pregnancy (associated most strongly with *PGR-TRPC6*, but also with *IGFBP7* and *GDF15*). Overall, the PheWAS and MoBa studies suggest risk loci may contribute to maternal weight and/or body size before and throughout gestation.

As HG has been associated with adverse pregnancy and offspring outcomes such as increased risk for pre-eclampsia and small for gestational age[6,36], we explored the association between the HG risk genes and pregnancy characteristics and outcomes (pre-eclampsia, gestational diabetes, pregnancy duration and placental weight) and offspring characteristics and outcomes (perinatal death, birth weight). Pregnancy and outcome characteristics of the MoBa cohort self-reported and/or reported in the National Medical Birth Registry of Norway (NMBR) with normal NVP and HG are summarized in Supplementary Tables 10–12. Maternal BMI was not significantly different at the beginning of pregnancy for those with NVP compared to those with HG, but those with HG were significantly younger ($P = 4.3 \times 10^{-4}$) and had significantly lower BMI at week 15 ($P = 2.3 \times 10^{-8}$) and at the end of pregnancy ($P = 2.2 \times 10^{-3}$). There were no significant differences in placental weight, birth weight or adverse outcomes other than significantly shorter pregnancies ($P = 4.1 \times 10^{-4}$) for those with HG (Supplementary Tables 10–12).

Among the risk loci identified in the meta-analysis included in MoBa, *PGR-TRPC6* was moderately associated with pre-eclampsia in the NMBR ($P = 1.3 \times 10^{-3}$) and *GDF15* with pre-eclampsia labeled as 'serious' in the NMBR ($P = 6.7 \times 10^{-3}$) (Supplementary Table 9). A low or high placental growth can be a risk factor for early- and late-onset pre-eclampsia, respectively[37]. Placental growth imbalance therefore makes a possible link between HG and pre-eclampsia, but none of the variants found in this study was associated with placental weight in MoBa. However, the *PGR-TRPC6* risk variant was associated with pregnancy duration ($P = 2.9 \times 10^{-5}$) (Supplementary Table 9). No

https://doi.org/10.1038/s41588-026-02564-4

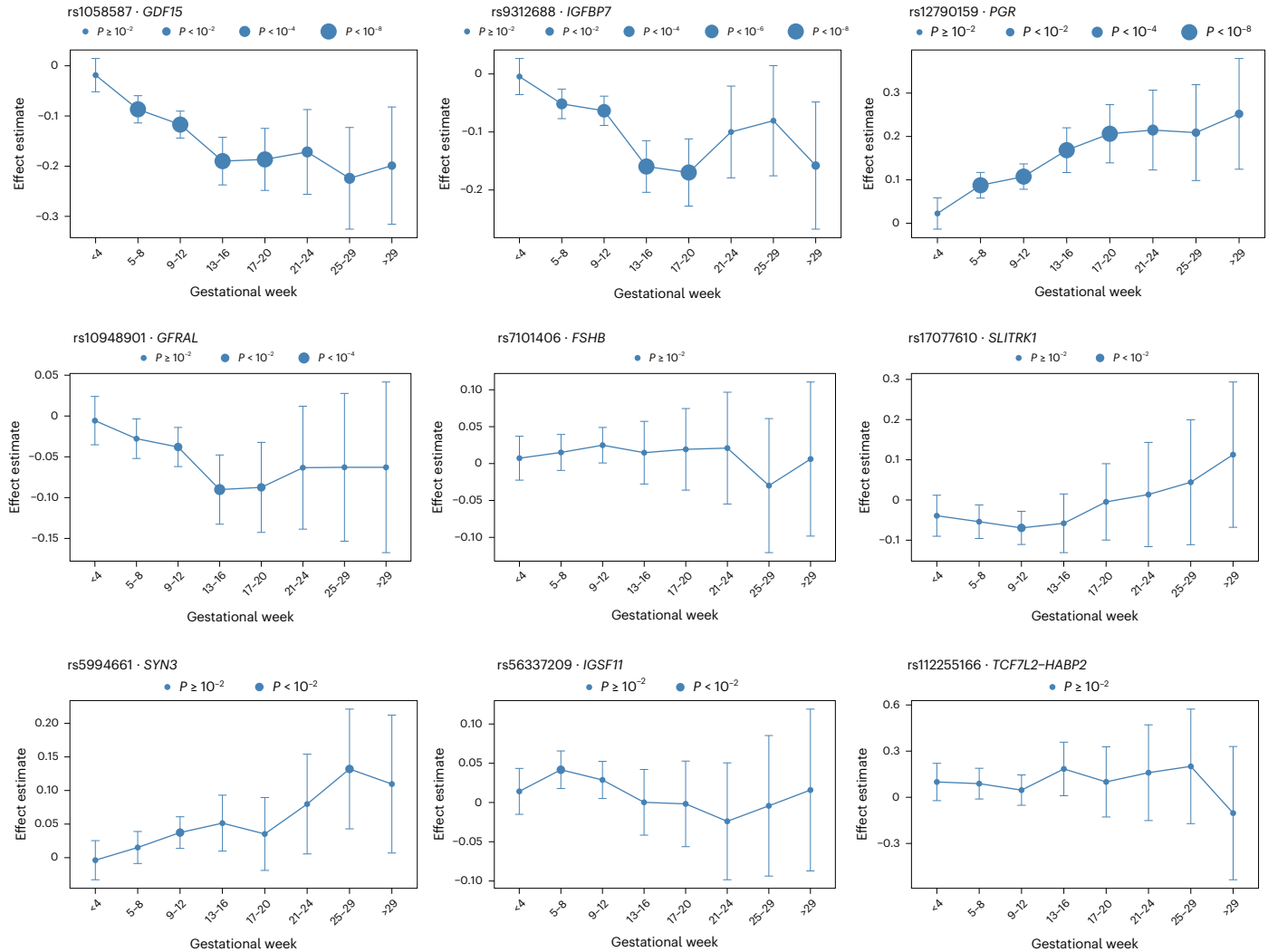


Fig. 4 | Effect-size estimates of lead SNPs associated with HG across gestational weeks in the MoBa cohort. For each lead SNP identified in the meta-analysis of GWAS, we examined the association of NVP status as reported by mothers against the number of alleles in the maternal genome, stratified by gestational week. The *y* axis represents the effect size estimates and their s.d. values at each timepoint, and the size of each point represents the significance level of the association. Points and error bars represent the effect size estimates and 95% CIs of the association with the maternal genome; $N = 56{,}225$ mothers. The values represented in the figure are available in Supplementary Tables 13–14.

association was identified between HG risk loci and perinatal death or gestational diabetes.

Temporal component to the genetic architecture of HG risk

Despite most NVP symptoms resolving by the end of the first trimester, HG lasts until term in 22% of cases[38]. We estimated effects of the risk loci on NVP during gestational ages <4, 5–8, 9–12, 13–16, 17–20, 21–24, 25–29 and >29 weeks' gestation in MoBa (Fig. 4 and Supplementary Table 13–14). We found that SNP effect sizes at loci *GDF15*, *PGR-TRPC6* and *IGFBP7* increased with time beyond the first trimester ($P < 0.002$), eventually stabilizing starting from weeks 13–16 until term ($P > 0.2$). The SNP at the *GFRAL* locus exhibited the largest effect on NVP at 13–16 weeks ($P = 0.001$). We note, however, that differential power over the cross-sectional analysis limits our ability to interpret effect size trajectories at these loci. Overall, we find evidence for a temporal component to the genetic architecture of HG risk.

Spatial trends of gene expression during placentation

Several of the genes associated with HG have been reported previously to be expressed during early placental development[10]. Therefore, using a spatial transcriptomics approach, we analyzed the expression of genes proximal to HG risk loci in fetal tissue and maternal tissue of the decidua during trophoblast differentiation, invasion and spiral artery remodeling. The highest levels of normalized expression in extravillous trophoblasts (EVT) of the fetal placenta and/or the spiral arteries of the developing maternal decidua were found for *GDF15*, *IGFBP7*, *TCF7L2* and *PGR* (Fig. 5). Consistent with our previous publication showing that most circulating GDF15 comes from the fetus during pregnancy[23], *GDF15* expression was not detected in the maternal spiral arteries and decidua. In the developing fetal placenta, the highest level of expression of *GDF15* was detected in the anchoring EVT (EVTA), with expression detected in floating villi (FV), villous cytotrophoblasts (VCT) and interstitial EVT (EVTI), but *GDF15* expression decreased significantly in endovascular EVT (EVTE) once EVT are inside the arteries. *TCF7L2* was expressed at high levels in EVTA with expression peaking in EVTI during decidual invasion, and lower levels detected in EVTE. Albeit at relatively lower levels, *TCF7L2* expression was detected in maternal decidua with low levels in spiral arteries. Conversely, expression of *IGFBP7* and *PGR* was limited to the maternal side with expression in spiral arteries declining with remodeling. Spatial trends and trends with artery remodeling during placentation were not detected for the remaining risk loci.

https://doi.org/10.1038/s41588-026-02564-4

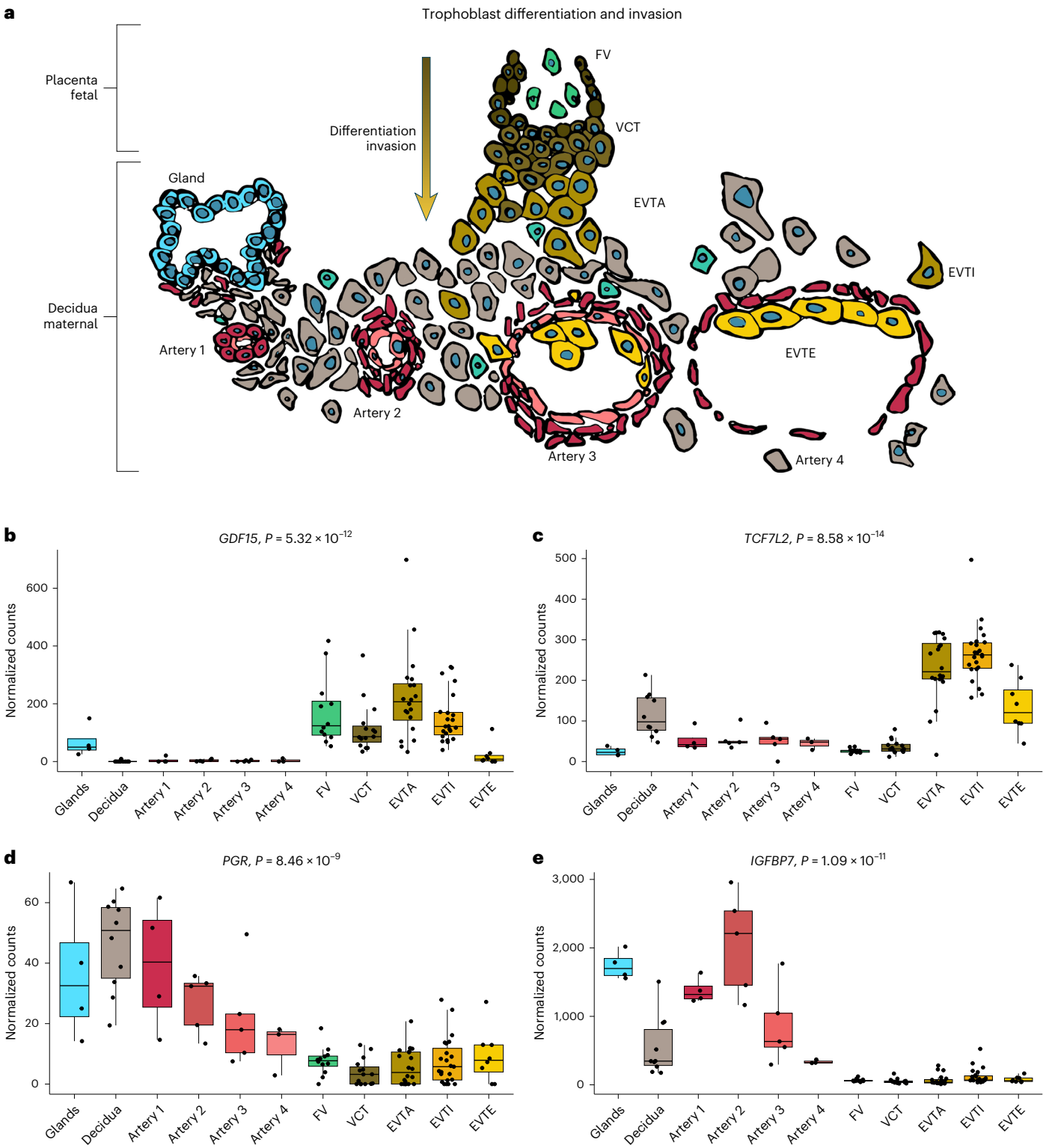


Fig. 5 | Spatial candidate gene expression during the first half of pregnancy. **a**, Model of the maternal-fetal interface during trophoblast differentiation, invasion and spiral artery remodeling. ROI were selected corresponding to one of the following categories: spiral arteries across four progressive stages of remodeling (Artery 1, 2, 3 and 4), annotated manually by expert ($N = 17$), decidua ($N = 10$), FV ($N = 12$), VCT ($N = 15$), EVTI ($N = 23$), EVTA ($N = 20$), EVTE ($N = 8$) and endometrial glands ($N = 4$) from 29 patients between 6 and 20 weeks' gestation. **b**–**e**, Normalized expression using NanoString whole transcriptome data for each ROI is shown for *GDF15* (**b**), *TCF7L2* (**c**), *PGR* (**d**) and *IGFBP7* (**e**). The boxplots display the 25th, 50th and 75th percentiles as the box, with whiskers extending to the most extreme values within 1.5× interquartile range.

Candidate gene expression in biologically relevant tissues

To determine whether candidate genes are expressed in biologically relevant tissues (female reproductive tissues (placenta, endometrium, ovary) and brain regions), mRNA expression datasets in Human Protein Atlas (proteinatlas.org)[39] were searched (Supplementary Table 15). Placental expression was detected for five of the ten risk genes from highest to lowest: *GDF15*, *IGFBP7*, *TCF7L2*, *PGR* and *SYN3*. Endometrial expression of IGFBP7, PGR, TCF7L2 and GDF15, and ovarian tissue

expression of *IGFBP7*, *TCF7L2*, *PGR*, *IGSF11* and *GDF15* was also observed. Furthermore, six candidate genes expressed in the brain were identified in this study. Among them, nearest genes *SYN3*, *SLITRK1* and *IGSF11* all have enriched (at least fourfold higher expression in brain compared to any other tissue analyzed) or enhanced (moderately elevated) expression in brain. Of note, *TCF7L2* (also known as *TCF4*) and *FSHB* are enriched specifically in the thalamus (brain region involved in feeding behavior) and the pituitary gland (brain region involved in reproduction), respectively.

Discussion

This GWAS meta-analysis of HG validated four known associations and identified six additional associations that map within, or in close proximity to, candidate genes *FSHB, SLITRK1, SYN3, IGSF11, TCF7L2* and *CDH9*. As historically the pregnancy hormones hCG and estrogen have both been hypothesized to be causal candidates for HG[1], it is important to emphasize that the meta-GWAS did not identify any association with the genes encoding these hormones or their receptors but rather identified an association with the gene encoding FSHB, and validated the strong association with the genes encoding the nausea and vomiting hormone (*GDF15*), its receptor (*GFRAL*) and the hormone receptor for progesterone (*PGR*). Future work linking hormones to HG should focus on these hormones, rather than continuing to expend limited resources on studies of hCG and estrogen given the lack of genetic evidence to support a connection. Of note, we found evidence for a regulatory element near *FSHB* in placenta in early pregnancy and FSH is expressed in the developing placenta with high expression in maternal decidua at 8 and 10 weeks' gestation, suggesting that, in addition to follicle stimulation, this hormone may play a role in placentation[40]. In addition, low FSH levels are a risk factor for chemotherapy-induced nausea and vomiting, suggesting there may be an inverse relationship between FSH and HG[41]. Another possible explanation for the association between FSHB and HG may be indirectly through increasing the risk of multiple gestation as the greatest risk locus for dizygotic twinning is in *FSHB* (rs11031005), which is in LD (r^2 = 0.2, D′ = 0.88) with the HG *FSHB* locus (rs7101406), and multiple gestation is an independent risk factor for HG[42,43].

Among the additional loci identified in this study are three that map within or near brain-enriched genes (*SYN3*, *SLITRK1* and *IGSF11*). *SLITRK1* and *IGSF11* did not show enrichment for placental expression, whereas *SYN3* showed evidence for expression in both placenta and brain. Although a role in placenta for SYN3 cannot be ruled out, we hypothesize these three neuronal synapse genes may play a role in the brain in downstream signaling of appetite, nausea and vomiting[44–46]. A role in appetite is supported indirectly by our finding in this study that *SLITRK1* is associated with prepregnancy maternal weight and BMI. Furthermore, *SYN3*, *SLITRK1* and *IGSF11* play roles in synaptogenesis and learning flexibility[45–47]. A notable feature of HG is heightened sensitivity to food, drink, sound and smell that leads to a learned association between these stimuli and nausea and vomiting. The mechanism known as conditioned taste aversion is associated with serotonin content and dopamine signaling[48,49] and *SLITRK1* knockout mice have serotonergic disturbances whereas *SYN3* regulates dopamine release[47,50]. Common antiemetic medications used to treat HG target serotonin and dopamine receptors and vary in effectiveness from one patient to the next, so future study to determine whether genotype-phenotype associations related to medication effectiveness in HG patients may be worthwhile[51].

Of the ten candidate genes in this study, six (*GDF15*, *GFRAL*, *IGFBP7*, *PGR*, *TCF7L2* and *SYN3*) have been associated with cachexia—a wasting condition with similar symptoms to HG including loss of appetite, weight loss and muscle wasting[10,52–57]. Manipulation of GDF15, GFRAL, IGFBP7, PGR and TCF7L2 in animal models has shown effectiveness in reducing symptoms of cachexia[53–57]. Thus, assuming analogous functions for these factors in HG, there is both genetic and biological support for causal and potentially reversible contributions for these genes in NVP.

The association between HG and six genes (*GDF15*, *GFRAL*, *TCF7L2*, *IGSF11*, *FSHB*, *IGFBP7*) that are related to insulin and altered glucose metabolism, provides biological support for the theory that, at least in some cases, glycemic status and insulin may be related to severity of NVP[58–63]. In addition, our PWAS identified ACP1—an enzyme that dephosphorylates the insulin receptor and plays a role in insulin sensitivity[64]. It has been postulated that reduced energy intake in early pregnancy results in lower maternal insulin and insulin-like growth factor 1, tipping the scales toward an evolutionary advantage that favors augmentation of placental development and subsequent fetal growth. This theory is supported by animal husbandry practices such as those designed to stimulate placental development by placing ewes on poor pasture land in early pregnancy[52]. However, first trimester protein deficiency in sows has been associated with more severe effects on fetal and placental growth than at any other time[65]. Insulin also plays a role in conditioned taste aversion, so another theory is that perhaps these risk loci abnormally dysregulate learning and memory formation of aversive stimuli[48].

Of note, a study of *TCF7L2* expression in mouse brain found neuronal expression in areas associated with sensing circulating nutrient levels including the area postrema[66]. The diabetes-associated gene is a transcription factor that may control glucagon-like peptide-1 (GLP-1) expression and is associated with liraglutide effects resulting in greater weight loss in obesity[67]. A link between TCF7L2 and GLP-1 is intriguing because the GLP-1 receptor (GLP1R) is expressed in the same cluster of neurons as GFRAL in the area postrema[67,68], and both GLP1R and GFRAL agonists are of great interest to pharmaceutical companies for their roles in appetite and weight loss[6,69]. Understanding where TCF7L2 fits into this unfolding landscape is therefore of particular interest.

In addition to neuronal expression, *TCF7L2* and three other candidate genes (*GDF15*, *IGFBP7* and *PGR*) were found in this study to be expressed differentially during early placental development. *GDF15* and *TCF7L2* were expressed primarily in fetal derived trophoblasts during differentiation and invasion. If these genes play roles in placental development, there is evidence for redundancy as *TCF7L2* knockout mice lack a placental phenotype and are developmentally normal at birth, and murine and human *GDF15* knockouts are reportedly normal and fertile[6,70–72]. Conversely, *IGFBP7* and *PGR* were found primarily in maternally derived spiral arteries, suggesting a maternal genetic contribution to spiral artery remodeling, and knockout of these genes in murine models contributes to a reduction in litter size and infertility, respectively[73,74]. In our transcriptome-wide association study, we identified an association with an antisense RNA to *IGFBP7*, suggesting a mechanism for altered expression of *IGFBP7*, despite the chr. 4 risk variant localization more than 350 kb from the closest gene (*IGFBP7*). In addition to the TWAS, the PWAS identified LILRB2, which is expressed in vascular smooth muscle placenta with hypothesized roles in placental immunity and extraembryonic vasculature[75].

This study also identified individual associations between risk genes and adverse outcomes including shorter pregnancy duration, pre-eclampsia and birth weight. HG has been associated with increased risk of pre-eclampsia[62] and the *PGR* risk locus for HG (rs12790159) is in LD (r^2 = 0.86, D′ = −0.95) with the risk locus for pre-eclampsia (rs2508372)[76]. Associations between increased GDF15 levels at 30–34 weeks' gestation and subsequent preterm pre-eclampsia have been noted previously[77], and herein a *GDF15* variant has been linked with serious pre-eclampsia. More studies and better statistical power are warranted to better understand the link between HG risk loci and adverse outcomes.

A principal purpose of genetic studies is to elucidate biological mechanisms associated with disease in the hopes that they will provide evidence to support the development of therapeutics that can benefit patients. The finding that genetic variation in *GDF15* is the most

significant association in this study and in a recent GWAS of Japanese HG patients[78], and the identification of a genome-wide significant variant 2.8 kb upstream of its receptor *GFRAL*, provide confidence in pursuing this therapeutic pathway. This study also predicted the *GDF15* risk locus to have opposing maternal and fetal contributions, consistent with our previous finding that lower circulating maternal GDF15 levels before pregnancy and higher circulating GDF15 levels produced by the fetal placenta during pregnancy are the main drivers of HG risk[23].

Finally, in this study we found highly disproportionate conservation of risk loci, which provides biological support for the theory that NVP evolved under strong selective pressure[62]. As approximately a third of pregnancies have no NVP symptoms, it may be an antiquated evolutionary mechanism no longer necessary for human survival. The ten genetic associations provide intriguing avenues to advance our understanding and pursue therapeutic pathways for a common pregnancy condition that in its most severe form is associated with substantial morbidity and even mortality for mothers and exposed offspring.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-026-02564-4.

References

1. Fejzo, M. S. et al. Nausea and vomiting of pregnancy and hyperemesis gravidarum. *Nat. Rev. Dis. Primers* **5**, 62 (2019).
2. Wang, H. et al. Severe nausea and vomiting in pregnancy: psychiatric and cognitive problems and brain structure in children. *BMC Med.* **18**, 228 (2020).
3. Gazmararian, J. Hospitalizations during pregnancy among managed care enrollees. *Obstet. Gynecol.* **100**, 94–100 (2002).
4. Nana, M. et al. Termination of wanted pregnancy and suicidal ideation in hyperemesis gravidarum: a mixed methods study. *Obstet. Med.* **15**, 180–184 (2022).
5. Munk-Olsen, T. et al. Postpartum depression: a developed and validated model predicting individual risk in new mothers. *Transl. Psychiatry* **12**, 419 (2022).
6. Fejzo, M. S. Hyperemesis gravidarum theories dispelled by recent research: a paradigm change for better care and outcomes. *Trends Mol. Med.* **30**, 530–540 (2024).
7. Almond, D., Edlund, L., Joffe, M. & Palme, M. An adaptive significance of morning sickness? Trivers–Willard and hyperemesis gravidarum. *Econ. Hum. Biol.* **21**, 167–171 (2016).
8. Sasso, E. B. et al. Marijuana use and perinatal outcomes in obstetric patients at a safety net hospital. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **266**, 36–41 (2021).
9. First, O. K. et al. Patterns of use and self-reported effectiveness of cannabis for hyperemesis gravidarum. *Geburtshilfe Frauenheilkd.* **82**, 517–527 (2022).
10. Fejzo, M. S. et al. Placenta and appetite genes GDF15 and IGFBP7 are associated with hyperemesis gravidarum. *Nat. Commun.* **9**, 1178 (2018).
11. Fejzo, M. S., MacGibbon, K. W., First, O., Quan, C. & Mullin, P. M. Whole-exome sequencing uncovers new variants in *GDF15* associated with hyperemesis gravidarum. *BJOG* **129**, 1845–1852 (2022).
12. Fejzo, M., MacGibbon, K. & Mullin, P. 905: Hormone receptor genes PGR and GFRAL linked to hyperemesis gravidarum. *Am. J. Obstet. Gynecol.* **220**, S585–S586 (2019).
13. Gottesman, O. et al. The Electronic Medical Records and Genomics (eMERGE) network: past, present, and future. *Genet. Med.* **15**, 761–771 (2013).
14. Leitsalu, L. et al. Cohort profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).
15. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
16. Finer, S. et al. Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *Int. J. Epidemiol.* **49**, 20–21i (2020).
17. *Access Results* (FinnGen, accessed 9 January 2023); https://www.finngen.fi/en/access_results
18. Brumpton, B. M. et al. The HUNT study: a population-based cohort for genetic research. *Cell Genom.* **2**, 100193 (2022).
19. Brandlistuen, R. E. et al. Cohort profile update: the Norwegian mother, father and child cohort (MoBa). *Int. J. Epidemiol.* **54**, 382–388 (2025).
20. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the 'Sum of Single Effects' model. *PLoS Genet.* **18**, e1010299 (2022).
21. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
22. EGG Consortium et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).
23. Fejzo, M. et al. GDF15 linked to maternal risk of nausea and vomiting during pregnancy. *Nature* **625**, 760–767 (2024).
24. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
25. Mansour, O. et al. Prescription medication use during pregnancy in the United States from 2011 to 2020: trends and safety evidence. *Am. J. Obstet. Gynecol.* **231**, 250 (2024).
26. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
27. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
28. Loesch, D. P. et al. Identification of plasma proteomic markers underlying polygenic risk of type 2 diabetes and related comorbidities. *Nat. Commun.* **16**, 2124 (2025).
29. Xu, Y. et al. An atlas of genetic scores to predict multi-omic traits. *Nature* **616**, 123–131 (2023).
30. ENCODE Project Consortium et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
31. Wang, M. et al. Single-nucleus multi-omic profiling of human placental syncytiotrophoblasts identifies cellular trajectories during pregnancy. *Nat. Genet.* **56**, 294–305 (2024).
32. Luo, Y. et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **48**, D882–D889 (2020).
33. ENCODE Project Consortium An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
34. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44**, D877–D881 (2016).
35. Buniello, A. et al. Open Targets Platform: facilitating therapeutic hypotheses building in drug discovery. *Nucleic Acids Res.* **53**, D1467–D1475 (2025).
36. Fiaschi, L., Nelson-Piercy, C., Gibson, J., Szatkowski, L. & Tata, L. J. Adverse maternal and birth outcomes in women admitted to hospital for hyperemesis gravidarum: a population-based cohort study. *Paediatr. Perinat. Epidemiol.* **32**, 40–51 (2018).

37. Beaumont, R. N. et al. Genome-wide association study of placental weight identifies distinct and shared genetic influences between placental and fetal growth. *Nat. Genet.* **55**, 1807–1819 (2023).
38. Fejzo, M. S. et al. Symptoms and pregnancy outcomes associated with extreme weight loss among women with hyperemesis gravidarum. *J. Womens Health (Larchmt)* **18**, 1981–1987 (2009).
39. Uhlén, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
40. Stilley, J. A. W. et al. FSH receptor (FSHR) expression in human extragonadal reproductive tissues and the developing placenta, and the impact of its deletion on pregnancy in mice. *Biol. Reprod.* **91**, 74 (2014).
41. Yokokawa, T. et al. Influence of menopause on chemotherapy-induced nausea and vomiting in highly emetogenic chemotherapy for breast cancer: a retrospective observational study. *Cancer Med.* **12**, 18745–18754 (2023).
42. Mbarek, H. et al. Genome-wide association study meta-analysis of dizygotic twinning illuminates genetic regulation of female fecundity. *Hum. Reprod.* **39**, 240–257 (2024).
43. Bailit, J. L. Hyperemesis gravidarium: epidemiologic findings from a large cohort. *Am. J. Obstet. Gynecol.* **193**, 811–814 (2005).
44. Michetti, C. et al. The knockout of synapsin II in mice impairs social behavior and functional connectivity generating an ASD-like phenotype. *Cereb. Cortex* **27**, 5014–5023 (2017).
45. Hatayama, M. & Aruga, J. Developmental control of noradrenergic system by SLITRK1 and its implications in the pathophysiology of neuropsychiatric disorders. *Front. Mol. Neurosci.* **15**, 1080739 (2023).
46. Jang, S. et al. Synaptic adhesion molecule IgSF11 regulates synaptic transmission and plasticity. *Nat. Neurosci.* **19**, 84–93 (2016).
47. Moore, A., Linden, J. & Jentsch, J. D. *Syn3*gene knockout negatively impacts aspects of reversal learning performance. *eNeuro* **8**, ENEURO.0251-21.2021 (2021).
48. Aonuma, H. et al. Effects of 5-HT and insulin on learning and memory formation in food-deprived snails. *Neurobiol. Learn. Mem.* **148**, 20–29 (2018).
49. McCutcheon, J. E., Ebner, S. R., Loriaux, A. L. & Roitman, M. F. Encoding of aversion by dopamine and the nucleus accumbens. *Front. Neurosci.* **6**, 137 (2012).
50. Hatayama, M. et al. SLITRK1-mediated noradrenergic projection suppression in the neonatal prefrontal cortex. *Commun. Biol.* **5**, 935 (2022).
51. Fejzo, M. S. et al. Antihistamines and other prognostic factors for adverse outcome in hyperemesis gravidarum. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **170**, 71–76 (2013).
52. Wan, E. S. et al. Genome-wide association analysis of body mass in chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.* **45**, 304–310 (2011).
53. Groarke, J. D. et al. Phase 2 study of the efficacy and safety of ponsegromab in patients with cancer cachexia: PROACC-1 study design. *J. Cachexia Sarcopenia Muscle* **15**, 1054–1061 (2024).
54. Suriben, R. et al. Antibody-mediated inhibition of GDF15–GFRAL activity reverses cancer cachexia in mice. *Nat. Med.* **26**, 1264–1270 (2020).
55. Lee, J., Ng, K. G.-L., Dombek, K. M., Eom, D. S. & Kwon, Y. V. Tumors overcome the action of the wasting factor ImpL2 by locally elevating Wnt/Wingless. *Proc. Natl Acad. Sci. USA* **118**, e2020120118 (2021).
56. Madeddu, C., Macciò, A., Panzone, F., Tanca, F. M. & Mantovani, G. Medroxyprogesterone acetate in the management of cancer cachexia. *Expert Opin. Pharmacother.* **10**, 1359–1366 (2009).
57. Leong, M. L., Karjalainen, K. & Ruedl, C. TCF7L2 is a master regulator of muscle wasting in severe cancer cachexia. Preprint at *Research Square* https://doi.org/10.21203/rs.3.rs-2033935/v1 (2022).
58. Wylde, S., Nwose, E. & Bwititi, P. Morning sickness in pregnancy: mini review of possible causes with proposal for monitoring by diagnostic methods. *Int. J. Reprod. Contracept. Obstet. Gynecol.* **261**, 267 (2016).
59. Zhou, Y. et al. TCF7L2 is a master regulator of insulin production and processing. *Hum. Mol. Genet.* **23**, 6419–6431 (2014).
60. Kim, H., Takegahara, N. & Choi, Y. IgSF11-mediated phosphorylation of pyruvate kinase M2 regulates osteoclast differentiation and prevents pathological bone loss. *Bone Res.* **11**, 17 (2023).
61. Cheng, Y. et al. Follicle-stimulating hormone orchestrates glucose-stimulated insulin secretion of pancreatic islets. *Nat. Commun.* **14**, 6991 (2023).
62. Liu, Y. et al. Serum IGFBP7 levels associate with insulin resistance and the risk of metabolic syndrome in a Chinese population. *Sci. Rep.* **5**, 10227 (2015).
63. Zhang, S.-Y., Danaei, Z., Bruce, K., Chiu, J. F. M. & Lam, T. K. T. Acute activation of GFRAL in the area postrema contributes to glucose regulation independent of weight. *Diabetes* **73**, 426–433 (2024).
64. Gloria-Bottini, F. et al. Body mass index and acid phosphatase locus 1 in diabetic disorders. *Acta Diabetol.* **47**, 139–143 (2010).
65. Huxley, R. Nausea and vomiting in early pregnancy: its role in placental development. *Obstet. Gynecol.* **95**, 779–782 (2000).
66. Lee, S., Lee, C. E., Elias, C. F. & Elmquist, J. K. Expression of the diabetes-associated gene TCF7L2 in adult mouse brain. *J. Comp. Neurol.* **517**, 925–939 (2009).
67. Maselli, D. et al. Effects of liraglutide on gastrointestinal functions and weight in obesity: a randomized clinical and pharmacogenomic trial. *Obesity (Silver Spring)* **30**, 1608–1620 (2022).
68. Dowsett, G. K. C. et al. A survey of the mouse hindbrain in the fed and fasted states using single-nucleus RNA sequencing. *Mol. Metab.* **53**, 101240 (2021).
69. Huang, X. et al. Gut hormone multi-agonists for the treatment of type 2 diabetes and obesity: advances and challenges. *J. Endocrinol.* **262**, e230404 (2024).
70. Marsh, B. & Blelloch, R. Single nuclei RNA-seq of mouse placental labyrinth development. *eLife* **9**, e60266 (2020).
71. Boj, S. F. et al. Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell* **151**, 1595–1607 (2012).
72. Mulcahy, M. C. et al. GDF15 knockout does not substantially impact perinatal body weight or neonatal outcomes in mice. *Endocrinology* **165**, bqae143 (2024).
73. Chatterjee, S. et al. Loss of Igfbp7 causes precocious involution in lactating mouse mammary gland. *PLoS ONE* **9**, e87858 (2014).
74. Toufaily, C. et al. Impaired LH surge amplitude in gonadotrope-specific progesterone receptor knockout mice. *J. Endocrinol.* **244**, 111–122 (2020).
75. McIntire, R. H. et al. Novel HLA-G-binding leukocyte immunoglobulin-like receptor (LILR) expression patterns in human placentas and umbilical cords. *Placenta* **29**, 631–638 (2008).
76. Honigberg, M. C. et al. Polygenic prediction of preeclampsia and gestational hypertension. *Nat. Med.* **29**, 1540–1549 (2023).
77. Wertaschnigg, D. et al. Second- and third-trimester serum levels of growth-differentiation factor-15 in prediction of pre-eclampsia. *Ultrasound Obstet. Gynecol.* **56**, 879–884 (2020).
78. Yonezawa, Y. et al. Genome-wide association study of nausea and vomiting during pregnancy in Japan: the TMM BirThree Cohort Study. *BMC Pregnancy Childbirth* **24**, 209 (2024).

79. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **3**, 731 (2018).
80. Ochoa, D. et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res.* **51**, D1353–D1359 (2023).

[1]Department of Population and Public Health Science, Center for Genetic Epidemiology, University of Southern California Keck School of Medicine, Los Angeles, CA, USA. [2]UCL EGA Institute for Women's Health, University College London, London, UK. [3]Wolfson Institute of Population Health, Queen Mary University of London, London, UK. [4]Institute of Genomics, Estonian Genome Centre, University of Tartu, Tartu, Estonia. [5]Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [6]Department of Public Health and Nursing, NTNU, HUNT Center for Molecular and Clinical Epidemiology, Norwegian University of Science and Technology, Trondheim, Norway. [7]Department of Public Health and Nursing, NTNU, HUNT Research Centre, Norwegian University of Science and Technology, Levanger, Norway. [8]Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [9]Division of Mental Health Care, St. Olavs Hospital, Trondheim, Norway. [10]Department of Clinical and Molecular Medicine, NTNU Norwegian University of Science and Technology, Trondheim, Norway. [11]Department of Research, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [12]Department of Obstetrics and Gynecology, Division of Quantitative and Clinical Sciences, Vanderbilt University Medical Center, Nashville, TN, USA. [13]Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA, USA. [14]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. [15]Department of Medicine, Division of Nephrology, Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, USA. [16]Hyperemesis Education and Research Foundation, Clackamas, OR, USA. [17]Key Laboratory of Multi-Cell Systems, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai, China. [18]Department of Pathology, Stanford University, Stanford, CA, USA. [19]Norwegian Institute of Public Health, Oslo, Norway. [20]Department of Clinical Science, Mohn Center for Diabetes Precision Medicine, University of Bergen, Bergen, Norway. [21]Department of Genetics and Bioinformatics, Health Data and Digitalization, Norwegian Institute of Public Health, Oslo, Norway. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: marlena.fejzo@med.usc.edu; marc.vaudel@uib.no; april.shu@med.usc.edu; nicholas.mancuso@med.usc.edu

Estonian Biobank Research Team

Andres Metspalu[4], Lili Milani[4], Tõnu Esko[4], Reedik Mägi[4], Mari Nelis[4] & Georgi Hudjashov[4]

Health Research Team

Eamonn Maher[22], Shabana Chaudhary[23], Joseph Gafton[23], Karen A. Hunt[23], Shapna Hussain[23], Kamrul Islam[23], Mohammed Bodrul Mazid[23], Elizabeth Owor[23], Jessry Russell[23], Nishat Safa[23], John Solly[23], Marie Spreckley[23], David A. van Heel[5], Jan Whalley[23], Ishevanhu Zengeya[23], Emily Mantle[23], Shaheen Akhtar[24], Samina Ashraf[24], Dan Mason[24], John Wright[24], Daniel MacArthur[25], Michael Simpson[26], Richard C. Trembath[26], Gerome Breen[26], Raymond Chung[26], Sang Hyuck Lee[26], Omar Asgar[27], Joanne Harvey[27], Karen Tricker[27], Caroline Winckley[27], Hanifa Khatun[27], Amna Asif[27], Claudia Langenberg[28], Grainne Colligan[29], Ceri Durham[29], Bill Newman[30], Ahsan Khan[31], Hilary Martin[32], Teng Heng[32], Matt Hurles[32], Vivek Iyer[32], Georgios Kalantzis[32], Vladimir Ovchinnikov[32], Iaroslav Popov[32], Klaudia Walter[32], Panos Deloukas[33], David Collier[33], Ana Angel[3], Saeed Bidi[3], Fabiola Eto[3], Sarah Finer[3], Chris Griffiths[3], Sam Hodgson[3], Benjamin M. Jacobs[3], Rohini Mathur[3], Caroline Morton[3], Asma Qureshi[3], Stuart Rison[3], Annum Salman[3], Miriam Samuel[3], Moneeza K. Siddiqui[3], Daniel Stow[3], Sabina Yasmin[3], Julia Zöllner[2,3] & Sheik Dowlut[3]

[22]Aston University, Birmingham, UK. [23]Blizard Institute, Queen Mary University of London, London, UK. [24]Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK. [25]Garvan Institute, Darlinghurst, New South Wales, Australia. [26]King's College London, London, UK. [27]Manchester University Hospitals, Manchester, UK. [28]Precision Healthcare University Research Institute, Queen Mary University of London, London, UK. [29]Social Action for Health (Charity), London, UK. [30]University of Manchester, Manchester, UK. [31]Waltham Forest Council, London, UK. [32]Wellcome Sanger Institute, Hinxton, UK. [33]William Harvey Research Institute, Queen Mary University of London, London, UK.

Methods

This research complies with all relevant ethical regulations with approval obtained for each of the nine datasets at their respective sites (23andMe (Ethical and Independent Review Services 23-110), HER (University of Southern California Institutional Review Board HS-20-00500), FinnGen (Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa HUS/990/2017), EstBB (Estonian Committee on Bioethics and Human Research 1.1-12/624), UKBB (North West Centre for Research Ethics Committee 11/NW/0382), Genes and Health (London South-East NRES Committee of the Health Research Authority (14/LO/1240), HUNT (Regional Committee for Ethics in Medical Research, Central Norway (2018/1622), eMERGE (Vanderbilt University Medical Center served as a central IRB IRB201609), MoBa (The Regional Committees for Medical and Health Research Ethics (West committee number 2012/67 and South-East committee B number 2014/1178)). Informed consent was obtained for all participants with no compensation.

Study cohorts and cohort-level GWAS

We included nine cohort-level GWASs (23andMe, HER ExWAS, FinnGen R8, EstBB, UKBB, Genes and Health, HUNT, eMERGE, MoBa) involving a total of 10,974 cases, 461,461 controls and 43,961,852 SNPs. The population make-up of each study is shown in Table 1. To the best of our knowledge, no repeated measurements were taken for the GWAS studies. All statistical tests were two-sided (where applicable). Although no sample size estimates were performed, our goal was to include >10,000 cases in our GWAS, over tenfold larger than the number of cases in our published epigenome-wide association study (which yielded significant findings), with the expectation that this would be sufficient to identify additional significant loci, should they exist. We did not perform replication in this study.

This is a study of a condition of pregnancy, so only people who had a pregnancy history with the condition were included as cases. Controls included a subset of participants defined as 'female.'

Biological males were not included in this study on a pregnancy condition. Case and control definitions by cohort are listed in Supplementary Table 1 with variation in criteria between cohorts. All cases and most controls excluded females with no pregnancy history (with the exception of HUNT). We used ancestries European, African, Asian (East and South Asian combined), Admixed American and Unknown based primarily on the Regeneron Genetics Center algorithm defined in the manuscript and used for the HER dataset. We used 'ethnicity' rather than 'ancestry' in the Genes and Health cohort description because we are describing the cohort in the terms of how we recruit—our related inclusion criteria is 'self-identifies as British-Bangladeshi or British-Pakistani ethnicity.' Once we have genetic data on our volunteers and can define genetic ancestry, we do so to ensure we are not conflating the social construct of ethnicity with genetic ancestry. Participants were not recruited for this study on aggregate data.

HER exome-wide association studies

Exome-wide summary statistics from an epigenome-wide association study on HG were used in this study[11]. Whole-exome sequencing was performed by Regeneron Genetics Center with the Illumina NovaSeq platform using paired-end 75-bp reads (Illumina). The captured bases were sequenced so that greater than 95% of samples passing initial QC had at least 90% of the targeted bases covered at 20× or greater. Paired-end reads and genetic variants were called using the Regeneron Genetics Center DNAseq analysis pipeline. Ancestry was assigned using standard genetic prediction methods based on the intersection of SNPs between HapMap3 and the nonfiltered project level variant call format. SNPs were filtered to include common, high-quality SNPs and merged for the HapMap3 dataset.

HER GWAS

We performed firth logistic regression with REGENIE[81] on 451,677 variants while adjusting for ten genotype principal component (PCs). We did not adjust for age because maternal age at time of affected/unaffected pregnancies could not be determined.

UKBB GWAS

We performed firth logistic regression with the REGENIE package on 59,366,184 variants while adjusting for age, genotype platform, and 40 genotype PCs. A total of 17,079,576 variants remained after removing variants with MAF < 0.001 and INFO score ≤ 0.3.

FinnGen GWAS

FinnGen is a public–private collaborative effort combining genotyping and digital health record data from Finnish health registries. We accessed publicly available summary statistics from release v.8 (refs. 17,82). Participants in FinnGen were genotyped with Illumina and Affymetrix chip arrays with QC to remove samples and variants of poor quality. Genome-wide imputation was performed using reference Finnish whole-genome sequence data. Sex, age, ten PCs and genotyping batch were included as covariates in the analysis. GWAS results for release v.8 were performed using REGENIE package. We filtered out association statistics for variants that exhibited an INFO score < 0.3 in any contributing substudy within FinnGen.

23andMe GWAS

Genome-wide summary statistics were provided from a GWAS on HG[10]. 23andMe, Inc. is a personal genetics company where customers can consent to participate as research subjects. Cases and controls were genotyped on one of four custom Illumina genotyping arrays, and additional genotypes were imputed using the September 2013 release of the 1000 Genomes Project Phase 133 reference haplotypes. All participants were filtered to select for European ancestry. Logistic regression was applied using age and five genotype PCs as covariates.

EstBB GWAS

The EstBB is a population-based biobank with more than 200,000 participants, representing 20% of the total Estonian population[83,84]. All EstBB participants were genotyped using Illumina arrays at the Core Genotyping Laboratory of the Institute of Genomics, University of Tartu. Samples were imputed using a population-specific imputation reference panel of 2,297 whole-genome sequencing samples[85]. REGENIE v.2.0.4 (ref. 81) was used for analysis, with age and the first ten PCs used as covariates.

Genes and Health GWAS

Genes and Health is a longitudinal population-based study of adult British-Bangladeshi and British-Pakistani people (SAS ethnicity) from East London, Bradford and Manchester. Data from the July 2021 data freeze was used, which includes 44,396 participants (see www.genesandhealth.org and the cohort profile)[16]. DNA was genotyped on the Illumina GSA3v chip and TOPMed-r2 imputed. REGENIE v.3.1.3 software was used to perform the GWAS using firth logistic regression (–firth approx).

HUNT GWAS

The Trøndelag Health Study (HUNT) is a population-based health survey including four recruitment waves—HUNT1 (1984–1986), HUNT2 (1995–1997), HUNT3 (2006–2008) and HUNT4 (2017–2019)—concentrated in the North-Trøndelag area, where all adults >20 years of age were invited to participate[86]. Electronic health records from the Trøndelag county hospitals (Nord-Trøndelag Hospital Trust, including St. Olavs, Namsos and Levanger Hospitals) hold International Classification of Diseases and Related Health Problems (ICD) codes back to 1987 and were last accessed 8 August 2021. DNA from 71,860 HUNT samples in HUNT2–3 used in this study were genotyped using one of three different Illumina HumanCoreExome arrays (HumanCoreExome12 v.1.0, HumanCoreExome12 v.1.1 and UM HUNT Biobank v.1.0)[18].

ICD codes were grouped into phecodes[87] ninth and tenth revision (ICD-9/ICD-10) codes indicating pregnancy and/or delivery. SAIGE v.0.43.1 was used with covariates birth year, genotyping batch and the first four PCs from genotypes. Variants with minor allele count <10 were excluded.

eMERGE GWAS

The eMERGE network is a national network combining DNA biorepositories with electronic health record systems[13]. It contains data from participants at more than ten sites across the USA. Participants from all eMERGE sites were genotyped using the Illumina 660WQuad array for those reported as European or unknown ancestry and the Illumina 1M-Duo array for those of African ancestry. Imputation and standard QC procedures performed on each site cohort from eMERGE, including filtering based on sample quality and composition (for example, sample genotyping call rate and relatedness) and marker quality (for example, marker genotype call rate, concordance and Hardy–Weinberg equilibrium), as well as additional QC procedures used to merge genotype data from each site (for example, strand orientation analysis)[88]. We performed firth logistic regression, adjusting ten genotype PCs, using PLINK[89]. Variants with info scores < 0.5 and MAFs < 0.001 were excluded.

MoBa GWAS

MoBa is a population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health (NIPH)[90]. Participants were recruited from Norway from 1999 to 2008. The cohort includes approximately 114,500 children, 95,200 mothers and 75,200 fathers. The NMBR is a national health registry containing information about all births in Norway.

Genotyping data were acquired collectively by several Norwegian research centers on various genotyping platforms and are managed by the NIPH, information on the different genotyping batches is available from the NIPH (github.com/folkehelseinstituttet/mobagen); this study was conducted on the v.1.5 of the genotypes. QC of the genotypes was conducted by the PsychGen team—details on the pipeline are available in the code repository (github.com/psychgen/MoBaPsychGen-QC-pipeline). The current study is based on v.10 of the participant questionnaire files. Pregnancies were classified as follows:

No NVP: the mother did not report nausea or vomiting at any point of the pregnancy

NVP: the mother reported nausea or vomiting at any time of pregnancy but did not report being hospitalized for it

HG: the mother reported being hospitalized due to prolonged nausea or vomiting

GWAS was conducted as a case–control HG versus No NVP using Regenie[81] v.3.2 as available from Bioconda[91] and operated using SnakeMake[92] using Saddlepoint approximation against the genome of the mothers.

GWAS meta-analysis

To enable coherent meta-analysis across all studies, we annotated datasets with Single Nucleotide Polymorphism Database release 155 (dbsnp155)[93], filtered out SNPs with MAF ≥ 0.001 and INFO score ≤ 0.3 and synchronized datasets based on genome build GRCh38 (ref. 94). We first lifted over 23andMe, HER, FinnGen, EstBB and UKBB from GRCh37 to GRCh38 using MungeSumstats[95]. Then, we meta-analyzed 43,961,852 SNPs in nine GWASs (23andMe, HER, FinnGen, EstBB, UKBB, Genes and Health, HUNT, eMERGE, MoBa) using fixed effect analysis with METAL[96] (v.2020-05-05) while allowing for heterogeneity (Supplementary Fig. 2). Furthermore, to ensure population homogeneity, we conducted a separate meta-analysis restricted to 38,769,878 SNPs in six European-only GWASs (FinnGen, 23andMe, EstBB, UKBB, HUNT, MoBa).

Fine mapping

We extracted SNPs present in both 515 unrelated Europeans participants from the 1000 Genomes Project and meta-analysis summary statistics. SNPs with inconsistent allele coding between the two datasets were flipped based on the 1000 Genomes Project. Ambiguous SNPs and SNPs with MAF below 0.001 were removed. The lead SNP on chr. 5, rs10073299, was dropped from this analysis, as it has an allele frequency of 0 in 1000 Genomes Europeans.

We performed two separate statistical fine mapping analyses using meta-analysis summary statistics from all nine multi-ancestry cohorts and from six European-only cohorts. We fine-mapped ten regions within 1-Mb distance from the ten lead SNPs in meta-analysis with minimum absolute correlation (minimum pairwise LD estimate) of 0.3 and coverage (posterior inclusion probability threshold) of 0.9 using susieR[20,97] (v.0.12.35)—a method that performs variable selection based on SuSiE (Sum of Single Effect) model and iterative Bayesian stepwise selection fitting procedure under a sparse Bayesian regression[20,21].

Enrichment analysis

We used LDSC (v.1.0.1) to perform two separate partitioned LD score regression[24] on meta-analysis summary statistics from all nine cohorts and from six European-only cohorts, using 1000Genomes[98] European Phase 3 baseline LD and HapMap3 (ref. 99) weights. Variants in the major histocompatibility complex regions have been removed due to the uncommon LD structure. We estimated heritability and enrichments for 97 functional annotations while converting heritability from observed scale to liability scale (2.3% sample prevalence and varying population prevalence)[1,24,25].

Annotations from scATAC from placenta tissues and ChIP–seq from brain tissues

We extracted scATAC–seq peaks from placental samples during early pregnancy[31] and plotted accessibility profiles within ±1 Mb of the meta-analysis lead SNPs.

Furthermore, we extracted ChIP–seq data for the histone modification H3K27ac from adult female tissues in ENCODE[30–33]. We accessed seven tissues: chorion (ENCFF740CCS), chorionic villus (ENCFF911IYQ), esophagus (ENCFF947UUA), placenta (ENCFF193DUP, ENCFF242AQE), trophoblast (ENCFF119OPO, ENCFF264TTT), progenitor cells (ENCFF362BJU, ENCFF992CAZ), and brain (NCFF281ZGJ, ENCFF390JSM, ENCFF745VDW) and plotted enrichment peaks within ±1 Mb of the meta-analysis lead SNPs.

Transcriptome-wide association study

We used FUSION to perform TWAS using meta-analysis summary statistics, LD reference from 1000 Genomes Europeans and precomputed expression weights from 27 GTEx v.8 tissues and 1 brain tissue from Gandal et al.[26]. Tissue-specific significance threshold is determined using Bonferroni correction, calculated as 0.05 divided by the number of genes tested in each tissue[21,26,27].

PWAS was conducted to identify proteins whose genetically predicted abundance is associated with HG. Using the *FUSION* framework, we integrated HG GWAS summary statistics with harmonized proteomics genetic scores from OmicsPred.org (refs. 27,29). The genetic scores were extracted on three large-scale proteomics datasets by two publications: Xu et al. (SomaScan v.3, OPD000001), Xu et al. (Olink, OPD000002) and UKBB Multi-ancestry (Olink, OPD000007)[28,29]. Harmonized scoring files in GRCh38 were used. After QC, we performed PWAS independently in each proteomics dataset using GWAS summary statistics, and results from *cis*-pQTLs and *trans*-pQTLs were combined using Stouffer's *Z*-test method. Bonferroni correction was applied separately to correct for multiple testing within each proteomics dataset.

Phenome-wide association study

We conducted a PheWAS on ten variants identified by meta-analysis and around 3,380 phenotypes from three studies on Open Target Genetics, including UKBB, FinnGen and GWAS Catalog[35,80,99]. The *P* value threshold after the Bonferroni correction is 1.48×10^{-6}.

Annotations for regulatory features in Haploreg

Each lead SNP was entered into Haploreg v.4.2 to identify predicted regulatory features using default settings[34].

Associations with selected pregnancy outcomes in MoBa

Phenotypes for maternal characteristics (maternal weight/weight gain/BMI at beginning of pregnancy, at 15 weeks' gestation and at term), pregnancy characteristics (pre-eclampsia, gestational diabetes, pregnancy duration and placental weight), and offspring characteristics and outcomes (perinatal death, birth weight) were obtained from the participant questionnaires and from the NMBR. Association of NVP/HG status with binary phenotypes was evaluated using Fisher's exact test, association with linear phenotypes was evaluated using Student's *t*-test. Statistical significance was defined as a *P* value adjusted for the 13 continuous and 9 binary maternal characteristics and outcomes (<0.05 of 22) listed in Supplementary Tables 11 and 12.

Association with the lead variants of the loci genome-wide significant in the meta-analysis was evaluated for each phenotype (pre-eclampsia, gestational diabetes, pregnancy duration, placental weight, perinatal death and birth weight) using logistic or linear regression for binary and numeric phenotypes, respectively. For each pregnancy, a risk score was computed using the effect size estimates of the lead variants in the meta-analysis and the genotypes of the mother. Association of the phenotypes with the risk scores was computed as for the variants. All analyses were conducted in R v.4.2.2.

Spatiotemporal candidate gene expression during early placental development

The study's retrospective cohort design featured decidua tissue samples from elective pregnancy terminations at a single outpatient clinic[100]. Thorough medical history was captured, and a board-certified gynecologist extracted specific patient details. The tissue management approach involved reviewing whole tissue sections to identify suitable samples containing decidual tissue and spiral arteries. Selected blocks were assembled into a Tissue Microarray used here. Artery remodeling stage was determined manually by an expert annotator (E.S.). The spatial transcriptomics experiment was performed using NanoString Technologies according to company manuals[100]. Sample collection was performed as indicated in the GeoMx DSP instrument user manual (MAN-10088-03). Slides were loaded into the GeoMx DSP instrument and scanned. For each tissue sample, we selected regions of interest (ROI) corresponding to one of the following categories: artery ($N = 17$), decidua ($N = 10$), floating villi ($N = 12$), VCT ($N = 15$), interstitial EVT ($N = 23$), anchoring EVT ($N = 20$), endovascular EVT ($N = 8$), endometrial glands ($N = 4$) from 29 patients between 6 and 20 weeks' gestation; Each ROI was collected into a single well in a 96-well plate (see Supplementary Table 16 for patient ID and meta data per sample). We performed library preparation per manufacturer instructions. Libraries were paired-end sequenced (2 × 75) on a NextSeq550 with up to 400 million total aligned reads. We gathered raw counts of each gene from each sample using the NanoString GeoMx NGS processing pipeline. We performed QC aligned with NanoString's data analysis manual and default parameters. We used the approach recommended by the manufacturer for GeoMx data background subtraction using negative control probe reads and normalization. For normalization, we divided counts of all genes. This entailed dividing counts of all genes within a sample by that sample's 75th percentile of expression, followed by a multiplication with identical scaling factors for all samples—determined by the geometric mean of all 75th percentiles. We compared expression levels of each risk gene in all samples of the same origin combined, independent of gestational age, compared to the expression level of each of expression in samples of other origins (for example, comparing overall expression of a gene in all EVTA samples combined to overall expression of that gene in all EVTI samples combined). *P* value calculation for differential expression between samples of each origin was performed using the Kruskal–Wallis test comparing across the different sample-origin categories shown in Supplementary Table 16.

Conditioned parental and fetal effect estimates in MoBa

To estimate conditioned fetal, maternal and paternal effect estimates, we conducted separate GWASs of NVP severity encoded as 0, 1 and 2 for no NVP, NVP and HG, against genomes of the children, mothers and fathers in MoBa. NVP status and GWAS analyses were conducted as in the maternal GWAS used in the meta-analysis. Then, for the lead SNPs that were genome-wide significant in the meta-analysis, we applied a weighted linear model to the summary statistics of the fetal, maternal and paternal GWASs[22]. Specifically, the model estimates partitioned fetal, maternal and paternal effects from GWAS summary statistics equivalent to conditioned effects that would otherwise be obtained from a joint model given by

$$\text{phenotype} \sim \text{child} + \text{mother} + \text{father} + \text{covariates}$$

where phenotype here refers to NVP status; child, mother and father the number of tested alleles in the genome of the child, mother and father, respectively; and covariates the covariates used in the GWAS. The estimates of partitioned effects ($\hat{\eta}_c, \hat{\eta}_m$ and $\hat{\eta}_f$) for the child, mother and father, respectively, estimated from the GWAS ($\hat{\beta}_c, \hat{\beta}_m, \hat{\beta}_f$), respectively, are:

$$\hat{\eta}_c = 2\hat{\beta}_c - \hat{\beta}_m - \hat{\beta}_f$$

$$\hat{\eta}_m = -\hat{\beta}_c + \frac{3}{2}\hat{\beta}_m + \frac{1}{2}\hat{\beta}_f$$

$$\hat{\eta}_f = -\hat{\beta}_c + \frac{1}{2}\hat{\beta}_m + \frac{3}{2}\hat{\beta}_f$$

the s.e. values for the effect estimates are then estimated as

$$\hat{\sigma}\eta_c^2 = 4\hat{\sigma}c^2 + \hat{\sigma}_m^2 + \hat{\sigma}_f^2 + 2\left(\hat{\rho}'\right)_{mf}\sqrt{\hat{\sigma}_m^2\hat{\sigma}^2} - 4\left(\hat{\rho}'\right)_{cm}\sqrt{(\hat{\sigma}_c^2\hat{\sigma}_m^2)} - 4\left(\hat{\rho}'\right)_{cf}\sqrt{(\hat{\sigma}_c^2\hat{\sigma}_f^2)}$$

$$\hat{\sigma}\eta_m^2 = \tfrac{9}{4}\hat{\sigma}_m^2 + \hat{\sigma}_c^2 + \tfrac{1}{4}\hat{\sigma}_f^2 - \left(\hat{\rho}'\right)_{cf}\sqrt{\hat{\sigma}_c^2\hat{\sigma}_f^2} - 3(\hat{\rho}')_{cm}\sqrt{(\hat{\sigma}_c^2\hat{\sigma}_m^2)} + \tfrac{3}{2}\left(\hat{\rho}'\right)_{mf}\sqrt{(\hat{\sigma}_m^2\hat{\sigma}_f^2)}$$

$$\hat{\sigma}\eta_f^2 = \tfrac{9}{4}\hat{\sigma}_f^2 + \hat{\sigma}_c^2 + \tfrac{1}{4}\hat{\sigma}_m^2 - (\hat{\rho}')_{cm}\sqrt{(\hat{\sigma}_c^2\hat{\sigma}_m^2)} - 3(\hat{\rho}')_{cf}\sqrt{(\hat{\sigma}_c^2\hat{\sigma}_f^2)} + \tfrac{3}{2}\left(\hat{\rho}'\right)_{mf}\sqrt{(\hat{\sigma}_m^2\hat{\sigma}_f^2)}$$

where $\hat{\rho}'_{ij}$ are coefficients accounting for sample overlap and correlation between $\hat{\beta}_i$ and $\hat{\beta}_j$ estimated from the intercept of a bivariate LD Score regression. *P* values are calculated using a *Z*-test, with test statistic $Z = \frac{\hat{\eta}_i}{\hat{\sigma}_i}$[37,101].

Among the ten association signals in the meta-analysis, eight were found in the MoBa quality controlled genotypes, and one proxy association (rs112255166 $r^2 = 1$ for rs76856932 nearest genes *TCF7L2-HABP2*) was identified among two missing SNPs using TopLD[102]. Thus nine SNPs were available for analysis.

Maternal effects of individual SNPs by week of pregnancy

Mothers in MoBa reported whether they have suffered from NVP at different weeks of pregnancy (fhi.no/en/studies/moba/for-forskere-artikler/questionnaires-from-moba). To obtain cross-sectional effect size estimates during pregnancy, we conducted GWAS at these different timepoints using NVP as binary phenotype. GWASs were conducted as in the maternal GWAS used in the meta-analysis.

Statistics and Reproducibility
No statistical method was used to predetermine sample size.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
The summary statistics are available via Zenodo at https://doi.org/10.5281/zenodo.18274564 (ref. 103). The full GWAS summary statistics for the 23andMe discovery dataset will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Datasets will be made available at no cost for academic use. Please visit https://research.23andme.com/dataset-access/ for more information and to apply to access the data. Upon 23andMe data availability approval, the corresponding authors will share the full summary statistics for the nine cohorts upon request. Data from the MoBa study and the NMBR used in this study are managed by the national health register holders in Norway (NIPH) and can be made available to researchers, provided approval from the Regional Committees for Medical and Health Research Ethics, compliance with the European Union General Data Protection Regulation and approval from the data owners. The consent given by participants does not open for storage of data on an individual level in repositories or journals. Researchers who want access to datasets for replication should apply through https://helsedata.no/en/. Access to datasets requires approval from The Regional Committee for Medical and Health Research Ethics in Norway and an agreement with MoBa. Genes and Health: deposition of individual-level whole data from the Genes and Health study is not possible due to the governance model of the cohort and the commitments made to participants. All individual-level data are held securely within the Genes and Health Trusted Research Environment, and direct export or public deposition of genomic data is not permitted to protect participant confidentiality. Instead, access is available to registered researchers worldwide by application to the Genes and Health Executive (https://www.genesandhealth.org/). Applications are reviewed on a rolling monthly basis, and approved researchers are granted access to the individual-level data within the Trusted Research Environment, where analyses can be conducted, and can request the data files used in this study from the corresponding authors. GWAS data precomputed for all available phenotypes are freely downloadable under a CC BY-SA license. This can be accessed using the Google Cloud SDK (gcloud CLI), following the instructions in the documentation, and retrieving the dataset at https://storage.googleapis.com/genesandhealth_publicdatasets/ or https://console.cloud.google.com/storage/browser/genesandhealth_publicdatasets/.

Code availability
The code used to analyze MoBa data along with its documentation is available via Zenodo at https://doi.org/10.5281/zenodo.18675025 (ref. 104).

References

81. Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet.* **53**, 1097–1103 (2021).
82. *Other Maternal Disorders Predominantly Related to Pregnancy* (Risteys, accessed 9 January 2023); https://risteys.finregistry.fi/endpoints/O15_PREG_OTHER_MAT_DISORD
83. Pujol-Gualdo, N. et al. Advancing our understanding of genetic risk factors and potential personalized strategies for pelvic organ prolapse. *Nat. Commun.* **13**, 3584 (2022).
84. Koel, M. et al. GWAS meta-analyses clarify the genetics of cervical phenotypes and inform risk stratification for cervical cancer. *Hum. Mol. Genet.* **32**, 2103–2116 (2023).
85. Mitt, M. et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.* **25**, 869–876 (2017).
86. Åsvold, B. O. et al. Cohort profile update: the HUNT study, Norway. *Int. J. Epidemiol.* **52**, e80–e91 (2023).
87. Wu, P. et al. Mapping ICD-10 and ICD-10-CM codes to phecodes: workflow development and initial evaluation. *JMIR Med. Inform.* **7**, e14325 (2019).
88. Zuvich, R. L. et al. Pitfalls of merging GWAS data: lessons learned in the eMERGE network and quality control procedures to maintain high data quality. *Genet. Epidemiol.* **35**, 887–898 (2011).
89. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
90. Magnus, P. et al. Cohort profile update: the Norwegian mother and child cohort study (MoBa). *Int. J. Epidemiol.* **45**, 382–388 (2016).
91. The Bioconda Team et al. Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* **15**, 475–476 (2018).
92. Mölder, F. et al. Sustainable data analysis with Snakemake. *F1000Res.* **10**, 33 (2021).
93. Sherry, S. T., Ward, M. & Sirotkin, K. DbSNP—database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* **9**, 677–679 (1999).
94. Schneider, V. A. et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864 (2017).
95. Murphy, A. E., Schilder, B. M. & Skene, N. G. MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics* **37**, 4593–4596 (2021).
96. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
97. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B Stat. Methodol.* **82**, 1273–1300 (2020).
98. Fairley, S., Lowy-Gallego, E., Perry, E. & Flicek, P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* **48**, D941–D947 (2020).
99. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
100. Greenbaum, S. et al. A spatially resolved timeline of the human maternal–fetal interface. *Nature* **619**, 595–605 (2023).
101. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
102. Huang, L. et al. TOP-LD: a tool to explore linkage disequilibrium with TOPMed whole-genome sequence data. *Am. J. Hum. Genet.* **109**, 1175–1181 (2022).
103. Fejzo, M., Vaudel, M., Shu, C. & Mancuso, N. Multi-ancestry GWAS of severe pregnancy nausea and vomiting identifies risk loci associated with appetite, insulin signaling, and brain plasticity. *Zenodo* https://doi.org/10.5281/zenodo.18274564 (2026).
104. Vaudel, M. mvaudel/hyperemesis_gravidarum: publication. *Zenodo* https://doi.org/10.5281/zenodo.18675025 (2026).

Acknowledgements
We acknowledge the participant and investigators of 23andMe, HER, FinnGen, EstBB, UKBB, Genes and Health, HUNT, eMERGE and

MoBa. N.P.-G. was supported by MATER Marie Sklodowska-Curie, which received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 813707. M.A. was supported by 5U54CA20997105, 5DP5OD01982205, 1R01CA24063801A1, 5R01AG06827902, 5UH3CA24663303, 5R01CA22952904, 1U24CA22430901, 5R01AG05791504 and 5R01AG05628705 from the National Institutes of Health (NIH), W81XWH2110143 from the Department of Defense (DOD) and other funding from the Bill and Melinda Gates Foundation, the Cancer Research Institute, the Parker Center for Cancer Immunotherapy and the Breast Cancer Research Foundation. I.A. is an awardee of the Weizmann Institute of Science–Israel National Postdoctoral Award Program for Advancing Women in Science. E.S. is supported by National Science Scholarship, Agency for Science, Technology, and Research (A*STAR), Singapore. Y.G. and G.G. are supported by Shanghai Municipal Science and Technology Major Project. Genes and Health has recently been core-funded by Wellcome (WT102627, WT210561), the Medical Research Council (UK) (M009017, MR/X009777/1, MR/X009920/1), Higher Education Funding Council for England Catalyst, Barts Charity (845/1796), Health Data Research UK (for London substantive site) and research delivery support from the NHS National Institute for Health Research Clinical Research Network (North Thames). We acknowledge the support of the National Institute for Health and Care Research Barts Biomedical Research Centre (NIHR203330); a delivery partnership of Barts Health NHS Trust, Queen Mary University of London, St George's University Hospitals NHS Foundation Trust and St George's University of London. Genes and Health has recently been funded by Alnylam Pharmaceuticals, Genomics PLC and a Life Sciences Industry Consortium of AstraZeneca PLC, Bristol-Myers Squibb Company, GlaxoSmithKline Research and Development Limited, Maze Therapeutics Inc, Merck Sharp and Dohme LLC, Novo Nordisk A/S, Pfizer Inc. and Takeda Development Centre Americas Inc. E.A.J. was supported by the NIH Building Interdisciplinary Research Career's in Women's Health career development program (K12AR084232 principal investigators: A. S. Major and D. R. Velez Edwards). B.B., L.B. and K.H. work in a research unit funded by the Liaison Committee for education, research and innovation in Central Norway and the Joint Research Committee between St. Olavs Hospital and the Faculty of Medicine and Health Sciences, NTNU. N.M. was supported in part by NIH under awards R01HG012133, R01CA258808, R01GM140287 and U54HG013243. This study received funding from the Research Council of Norway (301178) and the European Research Council (ERC, AuditeMatrem, 101171420) to M.V. Further acknowledgement can be found in Supplementary Note.

Author contributions

M.F. conceived of and supervised execution of the study, analyzed and interpreted results, and drafted the paper. N.M. and C.A.S. supervised the meta-analysis and contributed to study design and paper preparation. X.W. performed the meta-analysis and contributed to paper preparation. X.W., Q.T., A. Kim, S.G., Y.G. and G.G. performed the functional analyses. A. Khan contributed to eMERGE genetic data analysis and phenotype harmonization. K.M. helped with data collection. M.V. designed and conducted the analyses in MoBa, and contributed to paper writing. P.M., S.J. and P.R.N. handled ethical and data access processes and provided scientific expertise and feedback on analyses and on the paper. N.P.-G. performed association analyses for EstBB data. N.P.-G. and T.L. contributed to writing and reviewing the paper. The EstBB Research Team played roles in data collection, genotyping, QC and imputation. I.A. analyzed and interpreted NanoString data, E.S. performed NanoString experiments and M.A. supervised the study. Y.G. performed experiments and data analysis. G.G. reviewed the paper. J.Z. and D.A.v.H. were involved in the generation of the GWAS data for Genes and Health. S.F. helped supervise the Genes and Health work, and S.F. and D.A.v.H. supervised the collection of the Genes and Health data. E.A.J. performed analyses and reviewed and revised the paper. D.R.V.E., T.E. and J.N.H. reviewed and edited the paper. G.P.J. obtained funding, contributed to design, and reviewed and edited the paper. L.B. was lead analysist for HUNT. K.H. and B.B. are co-principal investigators for HUNT. B.B. is lead contact for HUNT. Y.L. contributed data (for example, phenotyping for eMERGE cohort that is used by this study) and manuscript review.

Competing interests

M.F. declares the existence of financial/nonfinancial competing interests: HER Foundation (voluntary board member and research director); Harmonia Healthcare (chief scientific officer, stock, paid consultant); N.G.M. Biosciences (stock, paid consultant); Foundation for Women's Health (voluntary board member); The Morning Sickness Clinic (voluntary chief scientific officer). The other authors declare no competing interests.

Additional information

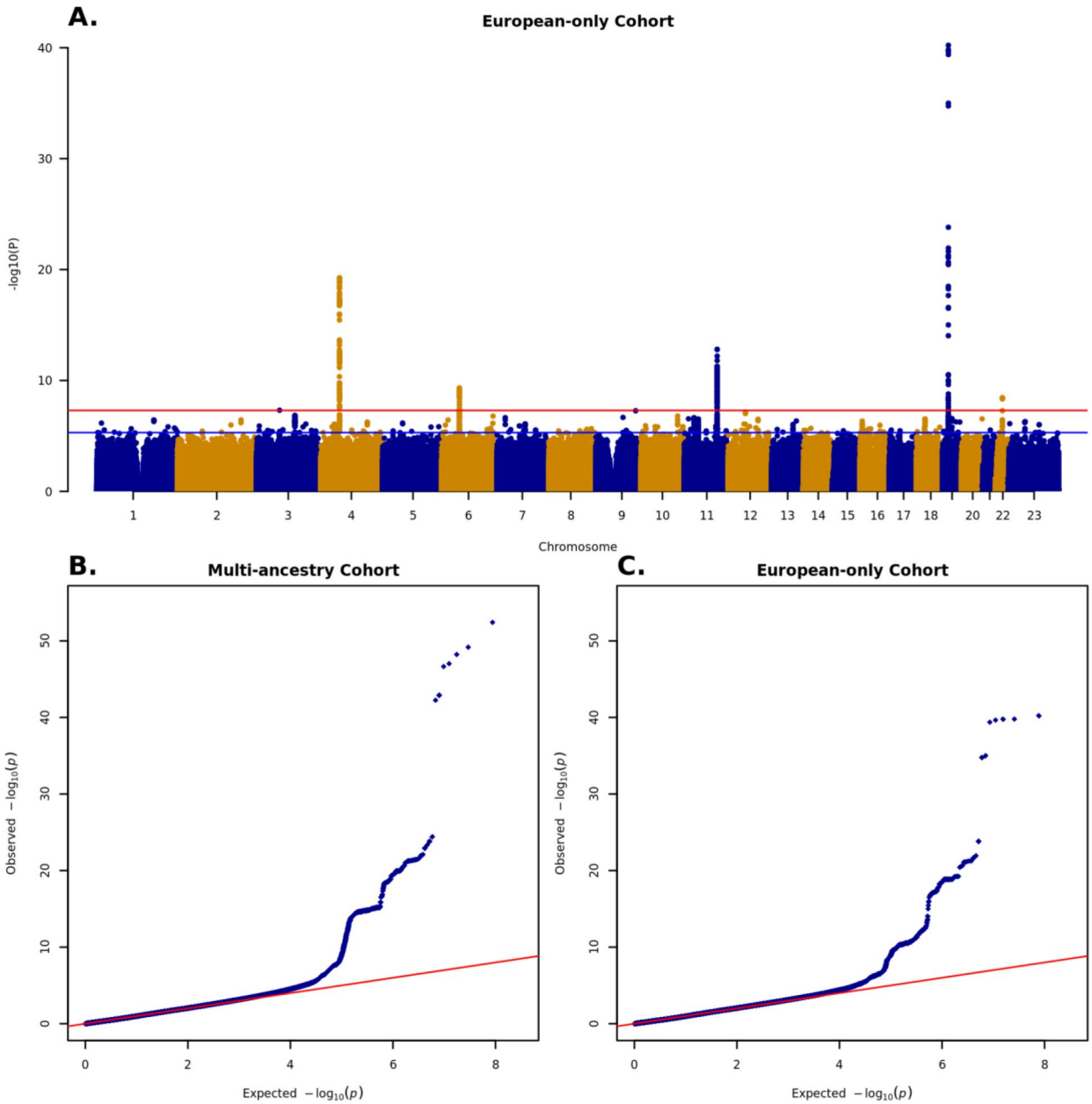
Extended data is available for this paper at https://doi.org/10.1038/s41588-026-02564-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-026-02564-4.

Correspondence and requests for materials should be addressed to Marlena Fejzo, Marc Vaudel, Chang April Shu or Nicholas Mancuso.

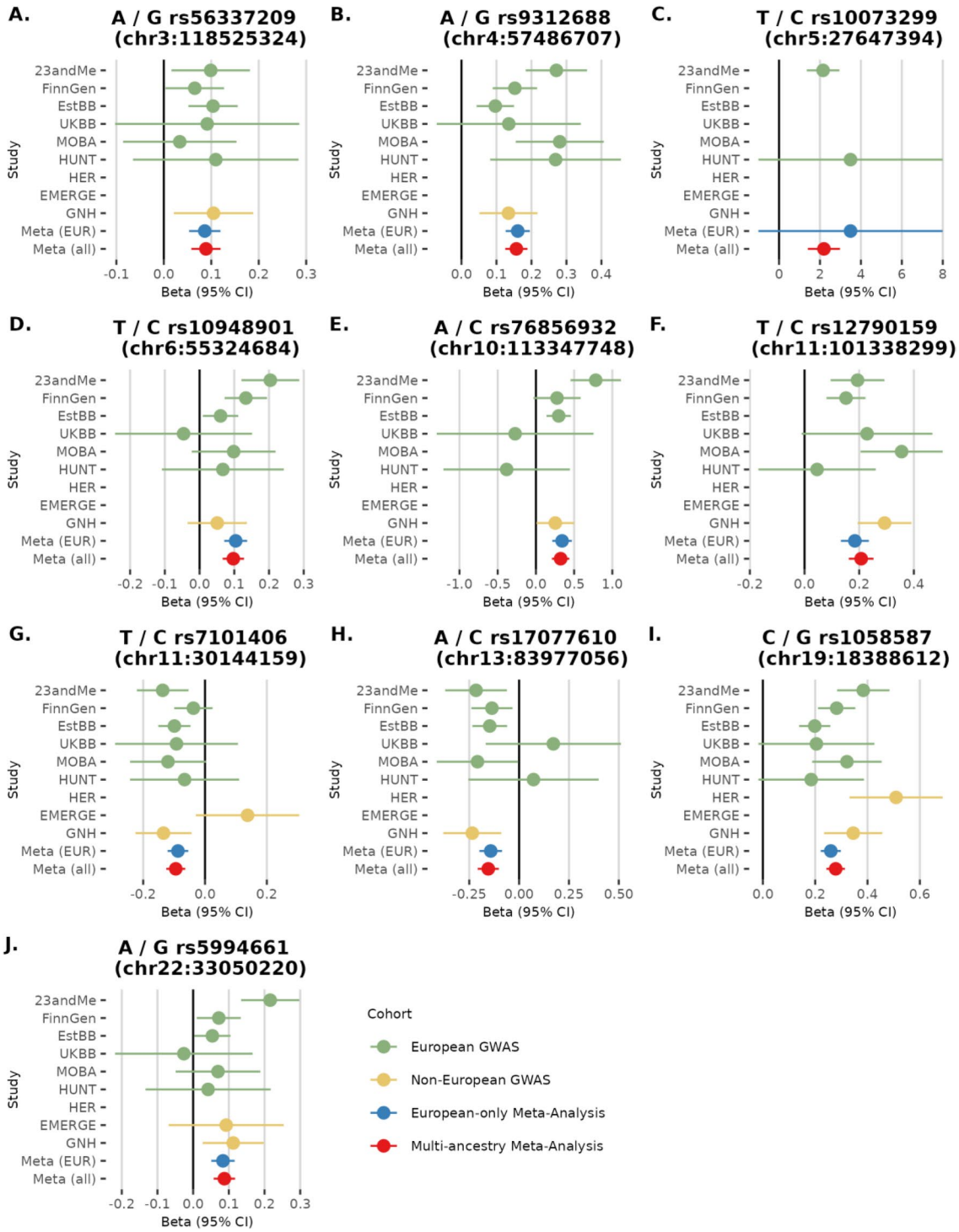
Peer review information *Nature Genetics* thanks Amanda Sferruzzi-Perri and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-026-02564-4



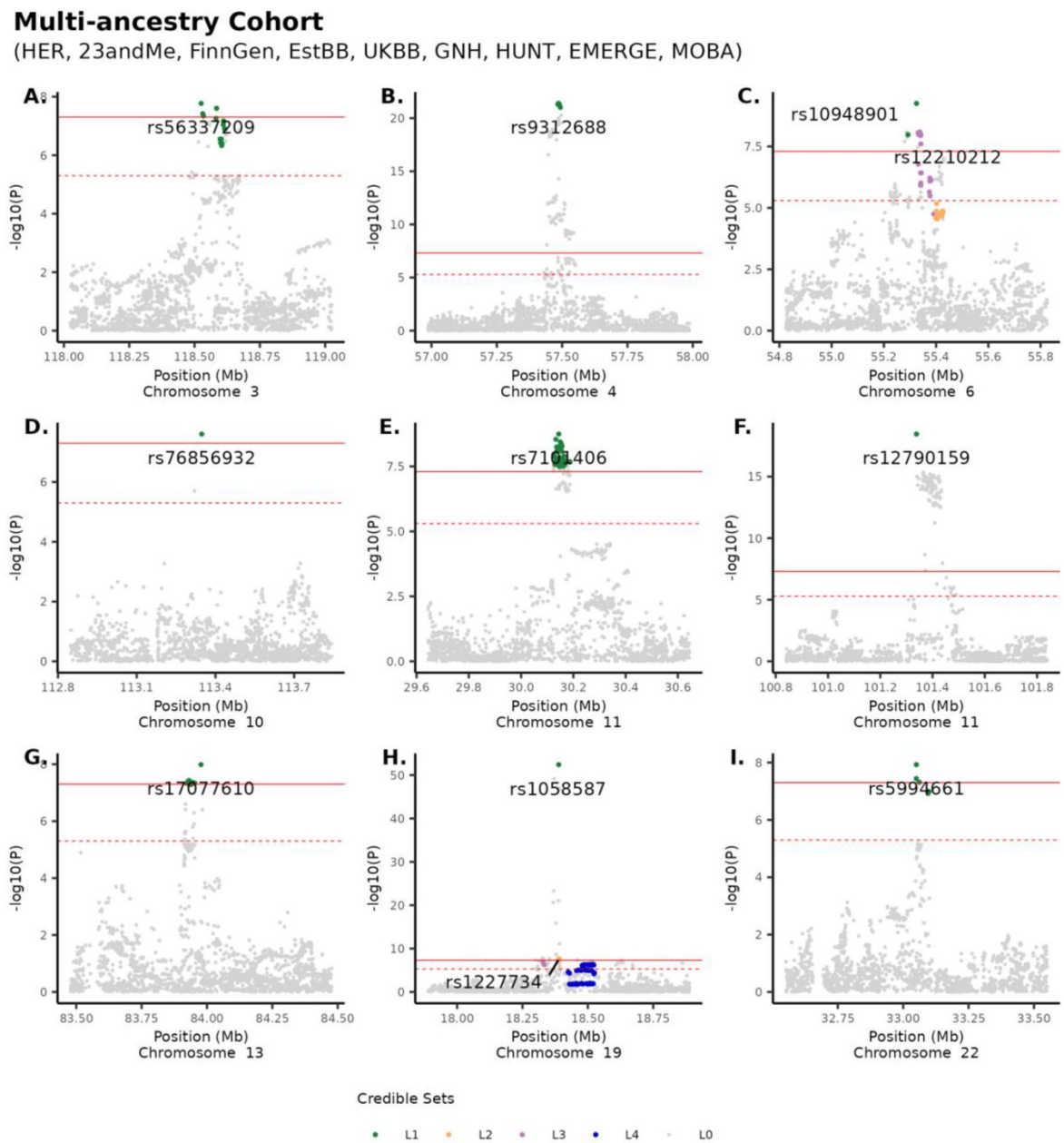
Extended Data Fig. 1 | Manhattan plot for European-only meta-analysis and QQ plots for cross-ancestry and European ancestry meta-analyses. A. Manhattan plot for European-only meta-analysis. X-axis represents genomic position, y-axis represents -$\log_{10}$(P). The red solid line represents the genome-wide significance threshold of $p < 5\times10^{-8}$ and the blue solid line represents the relaxed genome-wide significance threshold of $p < 5\times10^{-6}$. **B. QQ Plot for cross-ancestry meta-analysis. C. QQ plot for European-only meta-analysis**. X axis represents expected -$\log_{10}$(p), y axis represents observed -$\log_{10}$(p).

https://doi.org/10.1038/s41588-026-02564-4



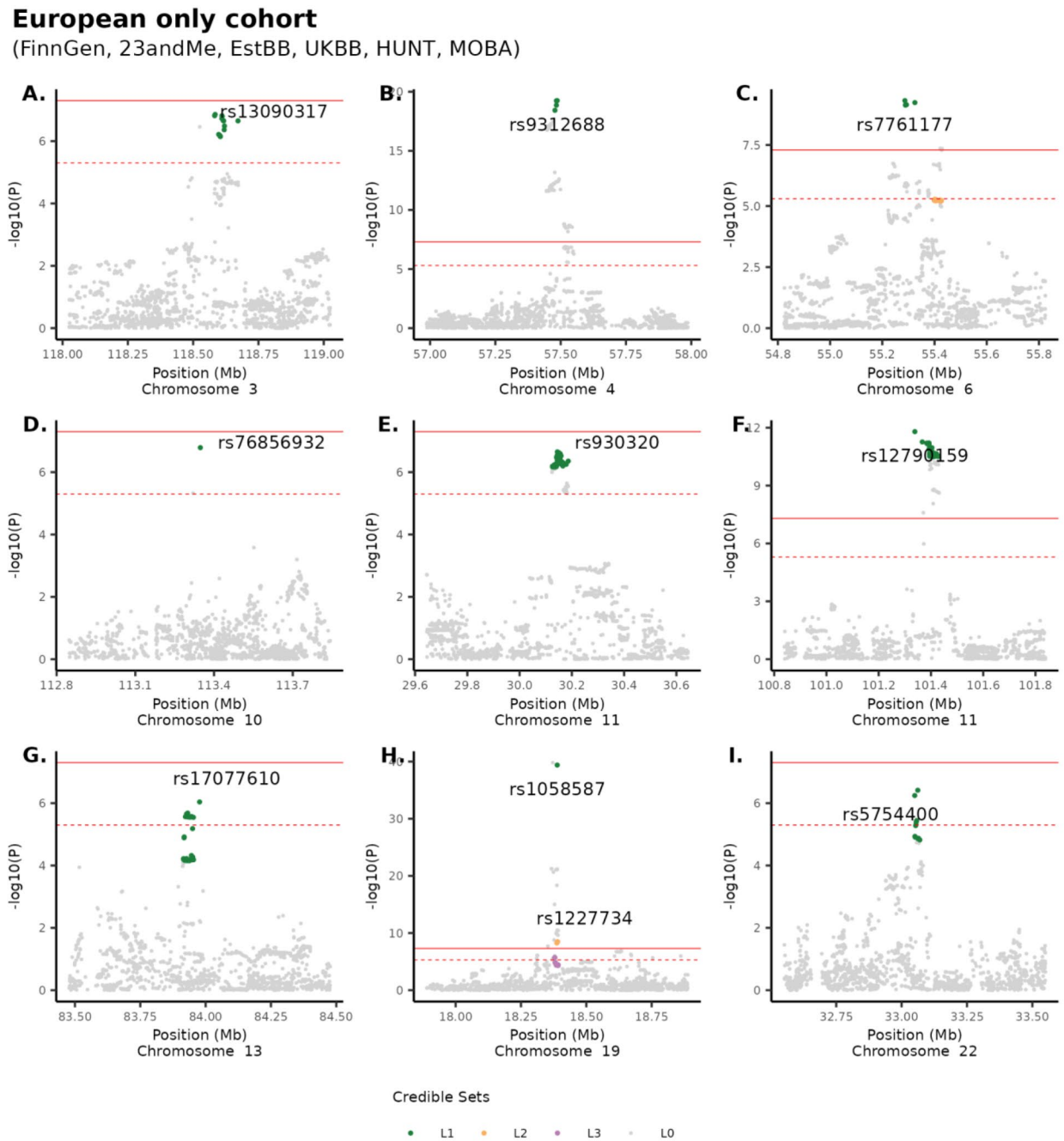
Extended Data Fig. 2 | Effect size estimates (95% confidence intervals) from individual GWAS datasets and meta-analyses. The y-axis lists results from nine individual GWAS cohorts and two meta-analyses. Green dots represent European GWASes, yellow dots represent non-European GWASes, blue dots represent the European-only meta-analysis, and red dots represent the multi-ethnic meta-analysis. **A.** rs56337209 **B.** rs9312688 **C.** rs10073299 **D.** rs10948901 **E.** rs76856932 **F.** rs7101406 **G.** rs12790159 **H.** rs17077610 **I.** rs1058587 **J.** rs5994661.

https://doi.org/10.1038/s41588-026-02564-4



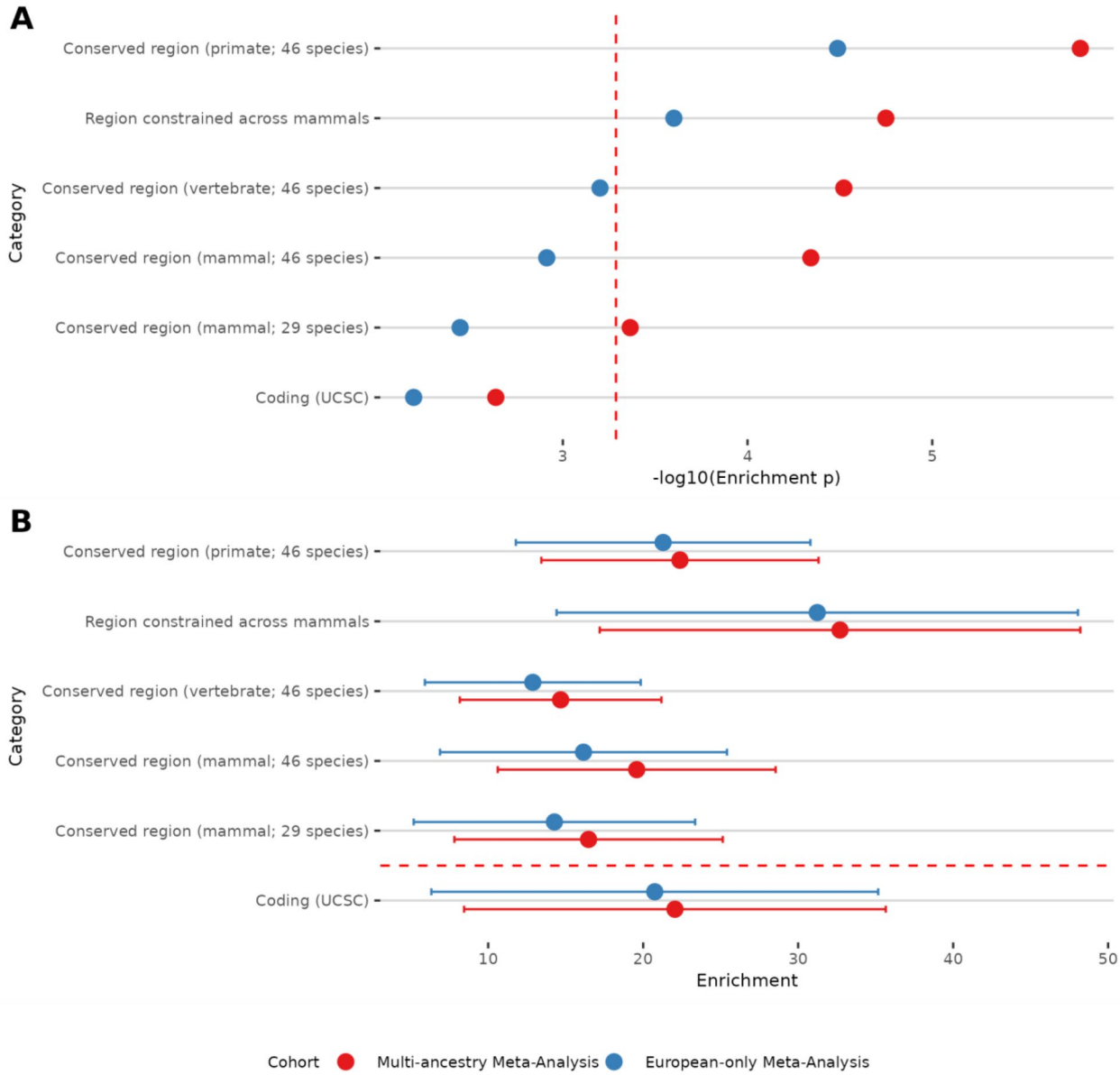
Extended Data Fig. 3 | Fine mapped signals for 9 lead SNPs in multi-ancestry meta-analysis. Chromosomal position (CRCh38) is shown in the X axis and -$\log_{10}$(P-value) is shown in the Y axis. Genome-wide significant threshold $p < 5\times10^{-8}$ is shown as a red dashed line and genome-wide significant threshold $p < 5\times10^{-6}$ is shown as red dotted line. SNPs not included in any credible set are highlighted in grey and SNPs in credible sets L1, L2, L3, L4 are highlighted in green, orange, pink, and blue. **A**. chr3:118.0-119.0 Mb **B**. chr4: 57.0-58.0 Mb **C**. chr6:54.8-55.8 Mb **D**. chr10:112.8-113.8 Mb **E**. chr11:29.6-34.6 Mb **F**. chr11:100.8-110.8 Mb **G**. chr13:83.5-84.5 Mb **H**. chr19:17.9-18.9 Mb **I**. chr22:32.5-33.5 Mb.

https://doi.org/10.1038/s41588-026-02564-4



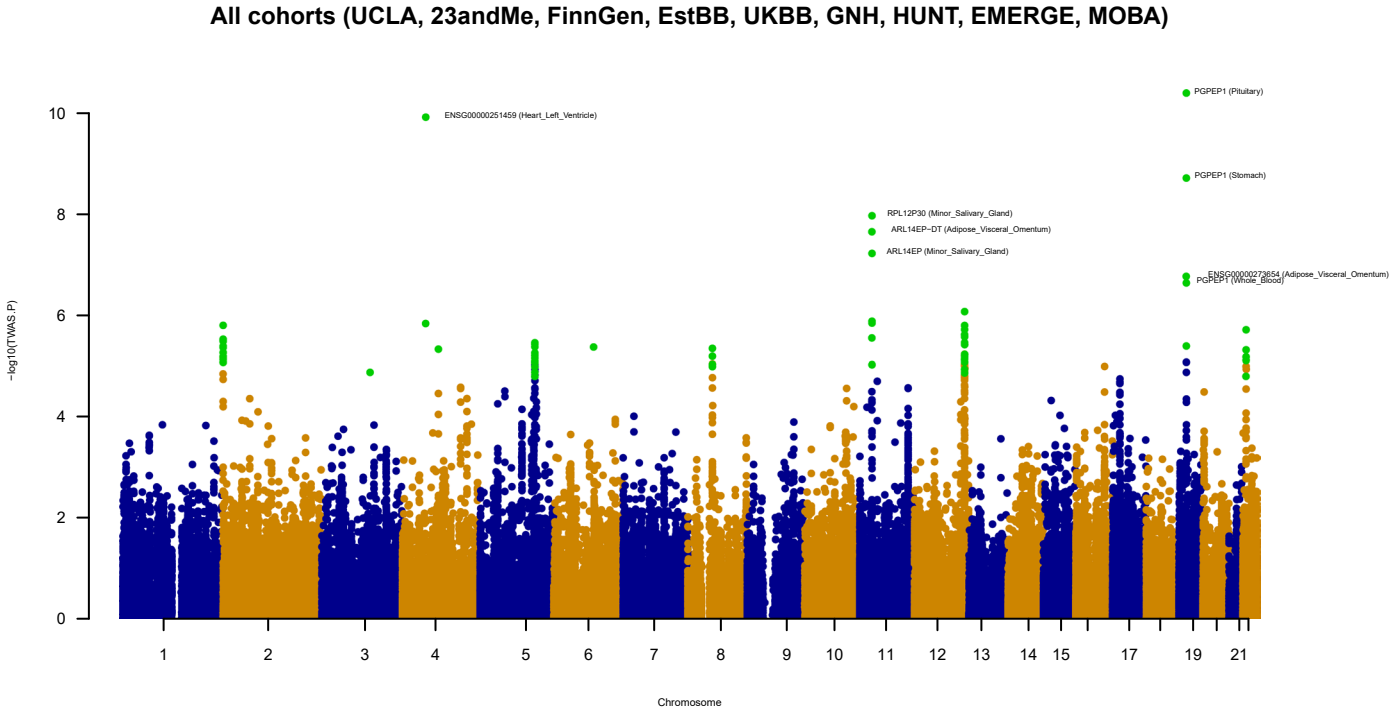
Extended Data Fig. 4 | Fine mapped signals for 9 lead SNPs in European-only meta-analysis. Chromosomal position (CRCh38) is shown in the X axis and -$\log_{10}$(P-value) is shown in the Y axis. Genome-wide significant threshold $p < 5\times10^{-8}$ is shown as a red dashed line and genome-wide significant threshold $p < 5\times10^{-6}$ is shown as red dotted line. SNPs not included in any credible set are highlighted in grey and SNPs in credible sets L1, L2, L3, L4 are highlighted in green, orange, pink, and blue. **A**. chr3:118.0-119.0 Mb **B**. chr4: 57.0-58.0 Mb **C**. chr6:54.8-55.8 Mb **D**. chr10:112.8-113.8 Mb **E**. chr11:29.6-34.6 Mb **F**. chr11:100.8-110.8 Mb **G**. chr13:83.5-84.5 Mb **H**. chr19:17.9-18.9 Mb **I**. chr22:32.5-33.5 Mb.

https://doi.org/10.1038/s41588-026-02564-4



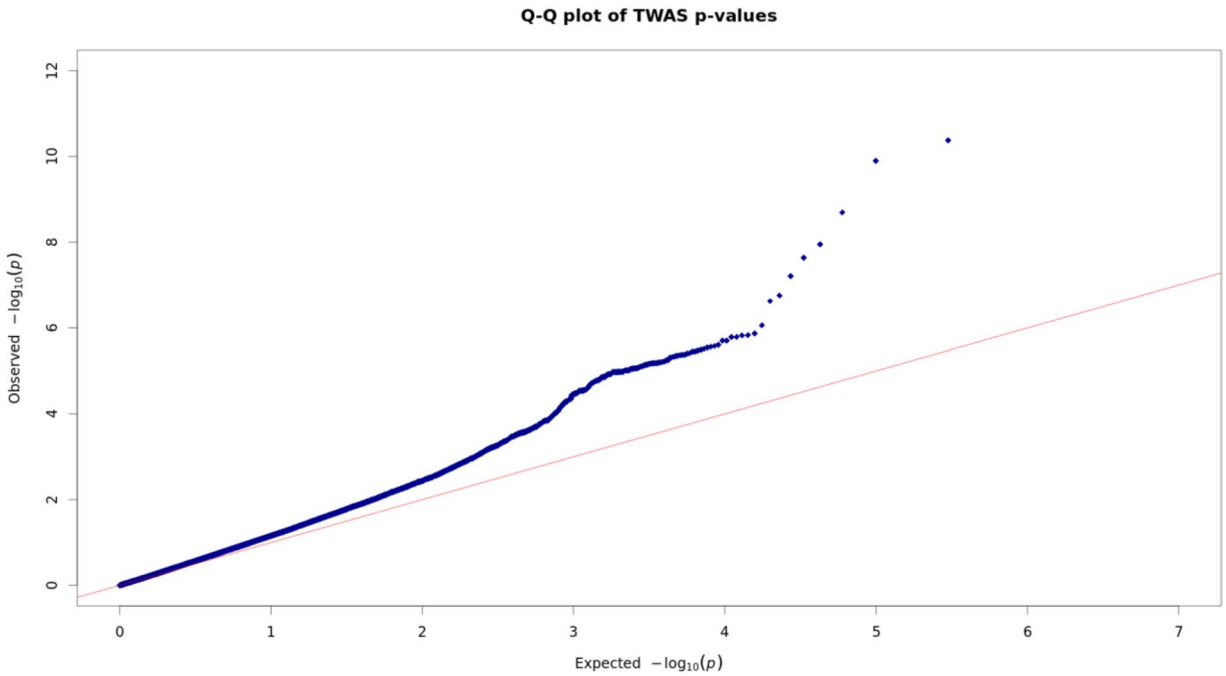
Extended Data Fig. 5 | Enrichment analysis suggests evolutionary constraint shapes genetic architecture of HG. A. Functional annotations by -log$_{10}$(Enrichment p). Functional annotation categories are shown in the Y axis and -log$_{10}$(Enrichment P-value) is shown in the X axis. **B. Functional annotation categories by Enrichment.** Functional annotation categories are shown in the Y axis and Enrichment (SE) is shown in the X axis. For both plots, P-value threshold after adjustment for multiple testing ($p < 0.0005$) is shown in the red dashed line. Heritability was converted from observed scale to liability scale (2.3% sample prevalence and 2% population prevalence). Blue dots represent European-only LDSC results, red dots represent multi-ethnic LDSC results.

https://doi.org/10.1038/s41588-026-02564-4



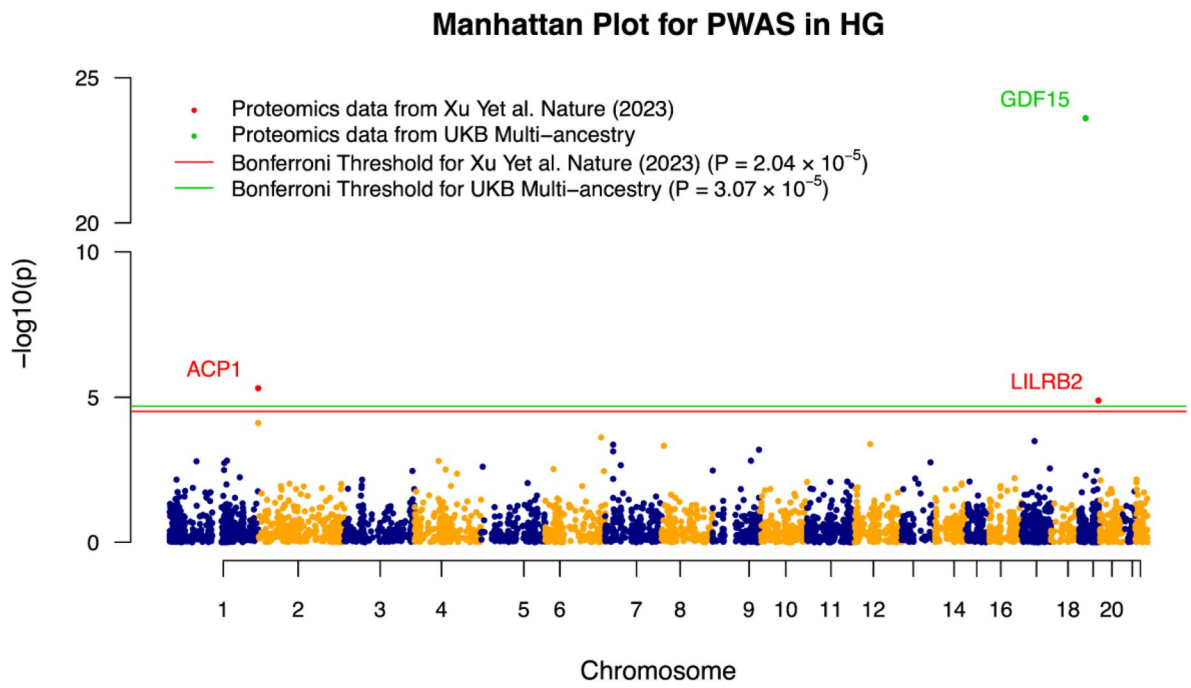
Extended Data Fig. 6 | Multi-ancestry transcriptome-wide association scan identified 65 significant gene-tissue pairs, representing 23 distinct genes. The Manhattan plot displays the distribution of TWAS test statistics across the genome, where -log10 transformed p-values are plotted against genomic positions for all 28 tissues (27 from GTEx v8 and one from Gandal et al.). Horizontal significance lines are not shown because tissue-specific significance thresholds were applied. Significant TWAS associations are highlighted in green. Due to the large number of significant signals (n = 65), individual gene annotations are omitted for clarity, except for signals surpassing significance threshold of p < 5e-6 (see Table 3).

https://doi.org/10.1038/s41588-026-02564-4



Extended Data Fig. 7 | QQ plot for multi-ancestry transcriptome-wide association scan for Hyperemesis Gravidarum. X axis represents expected -$\log_{10}$(p), y axis represents observed -$\log_{10}$(p).

https://doi.org/10.1038/s41588-026-02564-4



Extended Data Fig. 8 | Manhattan plot of proteome-wide association study (PWAS) results for HG across three proteomics datasets. PWAS were conducted from harmonised scores from three datasets OmicsPred.org: Xu Y. et al. *Nature* 2023 (SomaScan v3, OPD000001), Xu Y. et al. *Nature* 2023 (Olink, OPD000002), and UK Biobank Multi-ancestry (Olink, OPD000007). Bonferroni-corrected significance thresholds were applied separately for each dataset. Three proteins that passed the corresponding proteomics-wide thresholds were annotated. h^2_g.

nature portfolio

Corresponding author(s): Marlena Fejzo, Nicholas Mancuso, April Shu, Marc Vaudel

Last updated by author(s): 1.16.26

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	The code used to analyze MoBa data along with its documentation is available at https://doi.org/10.5281/zenodo.18675025. The rest of the analyses did not use any custom code or software.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The summary statistics are available at the following link: doi.org/10.5281/zenodo.18274564. The full GWAS summary statistics for the 23andMe discovery data set will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants.

Datasets will be made available at no cost for academic use. Please visit https://research.23andme.com/dataset-access/ for more information and to apply to access the data. Once the 23andMe data availability is approved, the corresponding authors will share the full summary statistics for the 9 cohorts upon request.
Data from the Norwegian Mother, Father and Child Cohort Study and the Medical Birth Registry of Norway used in this study are managed by the national health register holders in Norway (Norwegian Institute of public health) and can be made available to researchers, provided approval from the Regional Committees for Medical and Health Research Ethics (REC), compliance with the EU General Data Protection Regulation (GDPR) and approval from the data owners. The consent given by the participants does not open for storage of data on an individual level in repositories or journals. Researchers who want access to data sets for replication should apply through helsedata.no. Access to data sets requires approval from The Regional Committee for Medical and Health Research Ethics in Norway and an agreement with MoBa.
Genes & Health: Deposition of individual-level whole data from the Genes & Health study is not possible due to the governance model of the cohort and the commitments made to participants. All individual-level data are held securely within the Genes & Health Trusted Research Environment (TRE), and direct export or public deposition of genomic data is not permitted to protect participant confidentiality. Instead, access is available to registered researchers worldwide via application to the Genes & Health Executive (https://www.genesandhealth.org/). Applications are reviewed on a rolling monthly basis, and approved researchers are granted access to the individual-level data within the TRE, where analyses can be conducted and can request the data files used in this study from the corresponding author. GWAS data precomputed for all available pheno-types is freely downloadable under a CC BY-SA licence. This can be accessed using the Google Cloud SDK (gcloud CLI), following the instructions in the documentation, and retrieving the dataset at: gs://genesandhealth_publicdatasets/

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	This is a study of a condition of pregnancy, so only people who had a pregnancy history with the condition were included as cases. Controls included a subset of participants defined as "female."
Reporting on race, ethnicity, or other socially relevant groupings	We used ancestries European (EUR), African (AFR), Asian (East and South Asian combined), Admixed American (AMR), Unknown (UNK) based primarily on the Regeneron Genetics Center algorithm defined in the manuscript and used for the HER dataset. “Ancestry” is used throughout the manuscript with the exception the Genes & Health cohort, where an explanation for the use of “ethnicity” is specifically provided: We used this here because we are describing the cohort in the terms of how we recruit - our related inclusion criteria is "self-identifies as British Bangladeshi or British Pakistani ethnicity".
Population characteristics	Our discovery meta-analysis of GWAS on HG encompassed nine independent studies: three cohorts from the United States (23andMe, Inc., HER, and eMERGE), one from Estonia (Estonian Biobank),two from the United Kingdom (UK Biobank and Genes & Health), one from Finland (FinnGen Biobank, Release 8), and two from Norway (HUNTand MoBa). In total, we analyzed 10,974 cases and 461,461 controls across European (N=456,817), Asian (N=13,159), African (N=1,277), and Latino (N=75) ancestries (see Table 1). Broadly, cases were defined as patients diagnosed with HG or excessive vomiting during pregnancy, and controls were defined as females with no NVP diagnoses (see Supplementary Table 1). For each contributing study we performed rigorous Quality Control on genotyped and imputed variants, summary statistics, and outlier substudies (i.e., studies contributing to individual biobank results; see Methods).
Recruitment	Individuals were not recruited for this study on aggregate data.
Ethics oversight	Ethics approval was obtained for each of the 9 datasets at their respective sites.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Our goal was to include >10,000 cases in our GWAS, over 10-fold larger than the number of cases in our published EWAS.
Data exclusions	Males were not included in this study as it is a study on a pregnancy condition.
Replication	We did not perform replication in this study.
Randomization	NA
Blinding	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA