

ARTICLE OPEN



Endometriosis - on the intersection of modern environmental pollutants and ancient genetic regulatory variants

Amelia Warren¹, Demetra Andreou¹, Dean Warren², Jack Wieland¹ and Anna Mantzouratou¹✉

© The Author(s) 2025

Endometriosis is a chronic, estrogen-driven inflammatory disorder affecting approximately 10% of reproductive-aged women globally. Despite increasing genomic insights into advanced-stage disease, the genetic underpinnings of early-stage endometriosis remain poorly understood, limiting opportunities for timely diagnosis and intervention. This study explores the contribution of regulatory variants, including those derived from ancient hominin introgression, and their interaction with modern environmental exposures in shaping endometriosis susceptibility. We conducted a dual-phase literature review to identify genes implicated in endometriosis pathophysiology and endocrine-disrupting chemical (EDC) sensitivity. Five genes (*IL-6*, *CNR1*, *IDO1*, *TACR3*, and *KISS1R*) were selected based on tissue expression, pathway involvement, and EDC reactivity. Whole-genome sequencing (WGS) data from the Genomics England 100,000 Genomes Project were analysed in nineteen females with clinically confirmed endometriosis. Variant enrichment, co-localisation, and linkage disequilibrium analyses were conducted, and functional impact was evaluated using public regulatory databases. Six regulatory variants were significantly enriched in the endometriosis cohort compared to matched controls and the general Genomics England population. Notably, co-localised *IL-6* variants rs2069840 and rs34880821—located at a Neandertal-derived methylation site—demonstrated strong linkage disequilibrium and potential immune dysregulation. Variants in *CNR1* and *IDO1*, some of Denisovan origin, also showed significant associations. Several of these variants overlapped EDC-responsive regulatory regions, suggesting gene-environment interactions may exacerbate risk. These findings propose a novel perspective of endometriosis susceptibility, in which ancient regulatory variants and contemporary environmental exposures converge to modulate immune and inflammatory responses. This integrative approach identified new potential biomarkers for early-stage detection of endometriosis.

European Journal of Human Genetics; <https://doi.org/10.1038/s41431-025-01977-9>

INTRODUCTION

Globally, ten percent of reproductive-aged women have endometriosis, a heterogeneous gynaecological disease driven by estrogen signalling [1]. Endometriosis can be difficult to diagnose due to limited diagnostic tools, contributing to misdiagnosis and delays. Diagnosis can take up to eleven years between symptom onset and diagnosis [2]. Fifty percent of diagnosed women medically reported severe pelvic pain during adolescence that went untreated [3].

Endometriosis is potentially a multifactorial disease and may involve a complex system of immunological, environmental, hormonal, and genetic factors. Studies have suggested a dampened immune response in endometriosis patients due to estrogen dominance, triggering pro-inflammatory factors and altering immune cell functions. This fuels chronic inflammation and prevents cell death, promoting endometrial lesion growth [4]. Furthermore, a study using twins found a heritability component, with genome-wide association studies (GWAS) suggesting a genetic (47%) and environmental (53%) contribution to endometriosis predisposition [5]. Environmental predisposition from modern industrial pollutants and chemicals such as endocrine-disrupting chemicals (EDCs) may play a role in endometriosis

development. EDCs imitate hormones and block naturally occurring hormones from binding to receptors. This can interfere with physiological processes, including the reproductive system [6]. Current GWAS have collectively identified forty-two single nucleotide polymorphisms (SNPs) linked to endometriosis, some of which are associated with pain perception and maintenance and advanced endometriosis [7–9], however, none of these SNPs predict early endometriosis stages, hindering increased risk assessment accuracy and early diagnosis preventing complications like infertility.

Despite advancements in identifying endometriosis genes, research largely focuses on advanced stages and comorbidities, rather than disease onset, leading to a diagnosis of earlier stages and prevention of endometriosis to remain elusive. Understanding genetic risk and gene-environment interaction in early endometriosis is key to improving endometriosis management and preventing complications like infertility and gynaecologic cancers. This study aims to bridge the gap between genetics and environmental risk factors, providing a more comprehensive perspective of endometriosis susceptibility and identifying potential biomarkers for early-stage detection.

¹Department of Life and environmental Sciences, Bournemouth University, Bournemouth BH12 5BB, England. ²Independent Researcher, Poole, UK.
✉email: amantzouratou@bournemouth.ac.uk

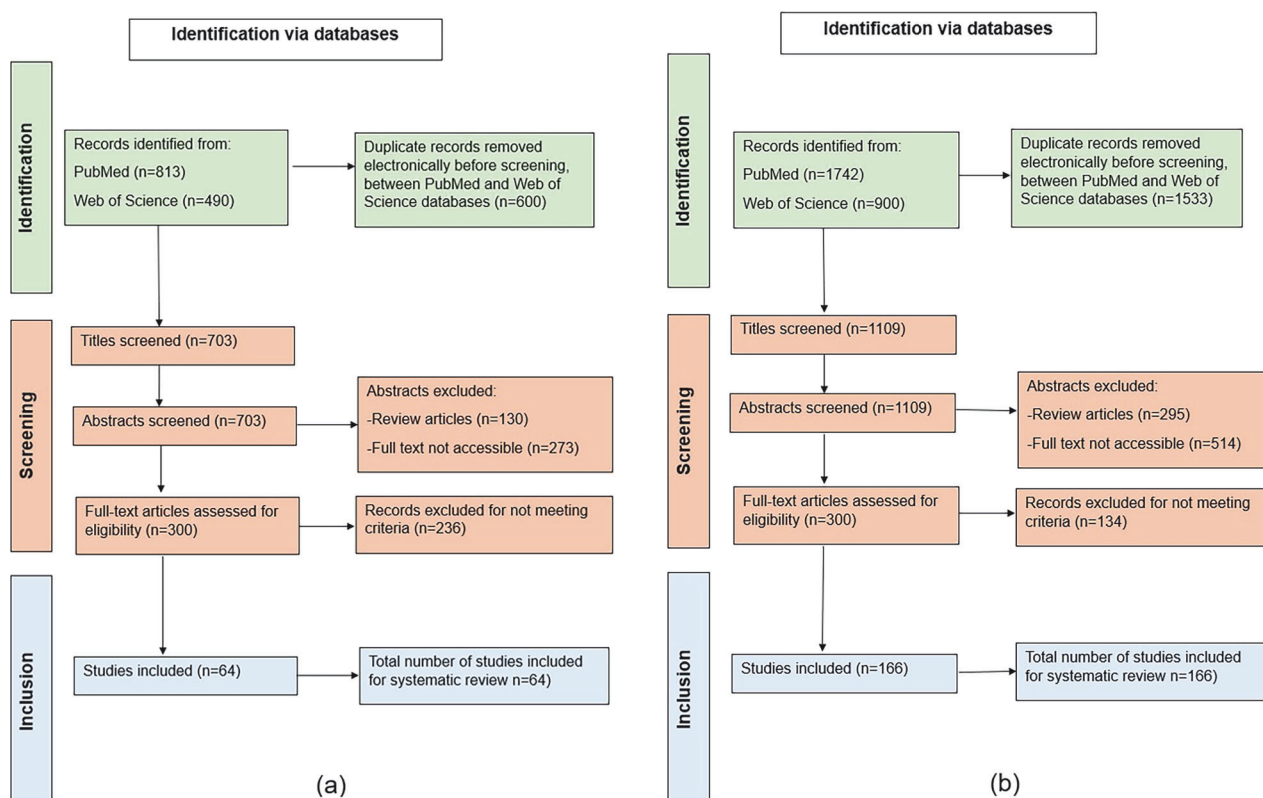


Fig. 1 PRISMA flow chart used for the screening process of literature searches. **a** PRISMA flow chart used to select literature for the impact of the environment on endometriosis risk **(b)** PRISMA flow chart created for the screening process to select literature on genomic areas of interest in relation to endometriosis development.

MATERIALS AND METHODS

Literature searches

A two-phase systematic literature review was conducted using PubMed and Web of Science to identify genes and genomic markers implicated in endometriosis pathophysiology and their interaction with environmental exposures, particularly EDCs.

Literature selection criteria. Inclusion Criteria: original studies focusing on genomic/genetic analysis, or genome-wide association design, only human participants (no other species), patients with a diagnosis of endometriosis for at least a year, patients aged between eighteen and forty-three years at the time of recruitment.

Exclusion Criteria: review studies, studies including participants with other types of female infertility, participants without an endometriosis diagnosis, participants with additional illnesses and diseases which could affect result outcomes.

A literature search investigating environmental risk factors for endometriosis and a review of their impact on signalling pathways and related genes was conducted. The key words used were: "endometriosis" and "exposure to endocrine disrupting chemicals", "endocrine disrupting chemicals", "exposure to pesticides", "pesticides", "personal care products", "cosmetics", "exposure to heavy metals", "heavy metals", "exposure to radiation", "radiation", "exposure to toxins", "toxins", "chemicals", "plastics", "exposure to pollution", "pollution", "exposure to air pollution", "air pollution", "exposure to water pollution", "water pollution". This yielded sixty-four papers and excluded 639 papers.

EDCs were prioritised based on scoping the literature corpus: 27/64 (42%) of included environmental studies evaluated EDCs.

A secondary literature search was conducted to identify genomic areas/markers of interest involved in endometriosis. The keywords used in this search were "endometriosis" and "polymorphism", "SNP", "genetic polymorphism", "variants", "locus", "GWA", "Genome-wide", "Genome wide", "Genetic association study". This yielded 166 papers and excluded 943 papers. Figure 1 shows the screening process of the literature searches as PRISMA flow charts.

From 57 candidates, five genes were pre-specified for a pilot, prioritising EDC responsiveness, pathway centrality, and expression at common implant

sites. We focused on non-coding regulation (introns, untranslated regions, promoter-flanking, ± 1 kb Transcription Start Site/Transcription End Site), given that EDCs more often perturb expression than protein sequence.

100,000 Genomics England database

The Genomics England (GE) 100,000 genome database [10, 11] was used as a part of an ongoing larger project "Genomic and chromosomal instability sequence markers in relation to fertility, early pregnancy, and cancers of the reproductive tissues" and included domains of Ovarian Cancer GE Research Network (GERN) and Endocrine and Metabolism GERN to obtain participants and analyse genomic data.

Participants from the GE Rare Disease Programs GRCh37 Participant Explorer were chosen by searching for a clinical diagnosis of endometriosis. All individuals were unrelated. No formal ancestry inference was undertaken in this pilot.

Inclusion Criteria: female participants aged eighteen to forty-three years at the time of recruitment, diagnosis of endometriosis recorded in medical history, availability of WGS data, inclusion of participants with endometriosis-related infertility and/or ovarian chocolate cysts.

Exclusion Criteria: individuals not assigned female sex at birth, participants over 43 years (to minimise confounding by menopause-related genetic changes), presence of additional ovarian pathology, chromosomal abnormalities, haematological disorders or other reproductive tract malignancies, diagnosis of diabetes, immunological disorders, or hormonal conditions that could confound genetic associations, body mass index outside the range of 18.5–30 kg/m² to minimize metabolic confounders' impact. The final study cohort included nineteen individuals meeting these criteria.

Each of the five selected genes were searched in GE per participant and single nucleotide variations (SNV), and insertion-deletion mutations (InDels) were collected.

This study focused on regulatory regions, introns, upstream, and downstream sequences, instead of coding regions, as environmental pollutants are more likely to affect gene expression than protein structure [12]. Targeting these regions enabled a more efficient and detailed investigation of relevant genomic variants.

Variant selection and filtering

Within the five pre-selected genes (*IDO1*, *IL-6*, *CNR1*, *TACR3*, *KISS1R*), candidate variants were extracted in the GE IVA workspace using Ensembl [13] variant effect predictor consequence categories corresponding to regulatory sequence. Variants present in the endometriosis cohort were prioritised if they overlapped regulatory annotations and mapped to pathways implicated in EDC response. This yielded ten candidate non-coding variants for testing.

Statistical analysis, linkage disequilibrium (LD) and Population branch statistic (PBS)

Heatmaps were created using GraphPad Prism v10 [14] to visualise variants (heatmaps are available on request). Variant frequencies were compared between the general GE population, and the endometriosis cohort, using R v4.2.2 a χ^2 goodness of fit test for individual variants, with Fisher's combined probability and small sample corrections relevant to data discovery. A Benjamini-Hochberg (BH) false discovery rate correction was applied to p-values, to account for multiple hypothesis testing controlling for false positives while maintaining statistical power (Fig. 2).

To confirm that variant enrichment was specific to the endometriosis cohort, nineteen randomly selected individuals without endometriosis from the GE database were screened using the same method. A Fisher's combined probability test compared variant frequencies across the endometriosis cohort, random group, and total GE population. Statistically significant variants were assessed for co-localisation to determine non-random clustering within the endometriosis cohort. LD analysis was conducted to assess the correlation between regulatory variants associated with endometriosis. For each pair of variants, we estimated the null probability of co-occurrence as the product of GE Total Population carrier proportions. We then compared the observed number of double-carriers using a one-sided tail test. Pairwise LD values (D' and r^2) were calculated for rs34880821 and rs2069840 in *IL-6* and rs806372 in *CNR1*, using data from the 1000 Genomes Project across multiple populations. LDpair and LDpop from LDlink [15–17] were used to determine linkage strength, with comparisons across African, East Asian, European, South Asian, and Admixed American populations. Results were analysed to evaluate population-specific evolutionary pressures and potential functional implications for immune regulation and pain sensitivity. We verified population allele frequencies for each rsID using Ensembl's Population Genetics (1000 Genomes Phase 3) [13] and gnomAD; for per-population AF alongside LD context we used LDlink. Within the GE Research Environment, we also extracted GE Total allele counts (allele count/allele number) via IVA to compute allele frequency used in our χ^2 and BH procedures. For rare variants like rs76129761, super-population minor allele frequency is very low, making r^2 estimates likely imprecise; accordingly, we did not present population LD for that SNP and focused *CNR1* LD reporting on rs806372, which has stable super-population metrics.

To contextualise population differentiation of candidate variants, we computed PBS for 1000 Genomes super-populations (AFR (African

population), EUR (European population), EAS (east Asian population)) using super-population allele frequencies from LDlink/1000G summaries. Pairwise F between populations was estimated from allele frequencies, transformed to branch lengths and combined. Negative PBS values were truncated to 0 for interpretation. PBS was calculated for rs76129761 and rs806372 (*CNR1*) and rs2069840 and rs34880821 (*IL-6*). Variants lacking reliable 1000G super-population coverage were not analysed.

Regulatory sequence analysis

To identify statistically significant variant effects a search of the variants rsID was undertaken using ClinVar, dbSNP, ensemble, UCSC, and String [13, 18–26]. An extensive analysis was conducted using UCSC [23] and previous literature on other hominoids and the conservation of the statistically significant variants. All results are reported with adjusted p-values, and statistically significant associations are discussed in the context of biological plausibility and environmental interactions.

RESULTS

Genomic findings of five selected genes and significant variant findings

Through using the literature search and consulting the criteria, the five genes chosen to be investigated further were *IDO1*, *IL-6*, *CNR1*, *TACR3*, and *KISS1R*.

Ten genetic variants were identified as potential contributors to endometriosis when compared to both the study population and the GE general population. Of these, 40% had a frequency of 0% in the Ensembl general population [13] but were present in the GE general population. This is potentially due to GE participants having a higher percentage of genetic variants through recruiting participants with known rare genetic conditions or specific diseases.

After conducting a χ^2 goodness of fit test for each of the ten variants found using the study cohort population frequency and GE total population frequency, variants, rs806372, rs76129761, rs2069840, rs34880821, rs933717388, and rs72643906 were found to be statistically significant when using a p value of 0.001. Following this a BH test was conducted, showing the variants are significant in the endometriosis cohort.

In the endometriosis cohort, 8/19 (47.4%) individuals carried both *IL-6* variants (rs2069840 and rs34880821). Using GE carrier proportions for these SNPs (6.24/19 and 4.80/19), the expected co-occurrence probability under independence 0.083, corresponding to 1.58 expected double-carriers in a nineteen-person sample. The observed count is therefore highly significant ($p = 7.4 \times 10^{-5}$). In the random nineteen-person group, 6/19 (31.6%) carried both,

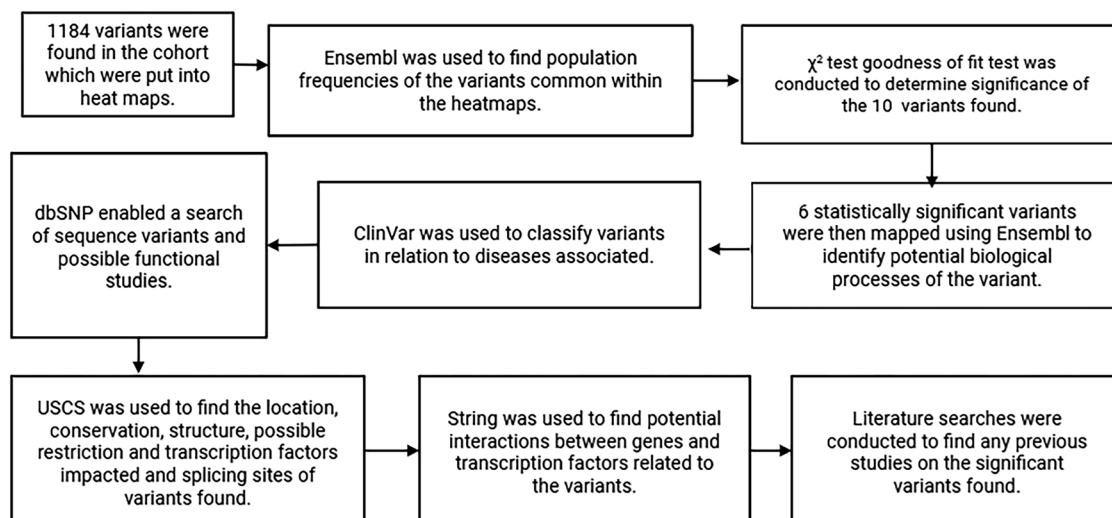


Fig. 2 Workflow of methods followed for finding the impact and regulatory sequences of statistically significant variants.

which is also higher than expected, though less pronounced ($p = 3.4 \times 10^{-3}$). This shows the co-localisation of rs2069840 and rs34880821 occurs much more frequently than would be expected by chance in both groups. Furthermore, both variants individually and together show significant enrichment in the endometriosis group. The elevated within-individual co-occurrence in random controls likely reflects underlying haplotype structure. Independence based expectations can underestimate co-occurrence when variants are correlated. We therefore treat co-occurrence qualitatively and emphasise case-control frequency contrasts for disease specificity. This suggests a potential biological interaction between these variants that may be particularly relevant to endometriosis.

When comparing the six significant variants in the endometriosis cohort compared to the random sample cohort and total GE populations, each of the variants is shown to be significantly higher in the endometriosis group, except rs76129761, which is shown to be higher in the random sample cohort. For rs76129761 (*CNR1*), counts were identical in cases and random controls (3/19 vs. 3/19). Both cohorts were nominally enriched versus the GE total population ($\chi^2 p = 0.023$ in each), after correction only the endometriosis cohort remained significant (BH $p = 0.046$ vs. 0.092 in the random cohort). We therefore treat rs76129761 as a context-dependent candidate rather than protective, and we avoid interpreting random-cohort carriers as evidence against association. From the Fisher's combined probability test between the endometriosis cohort and the random sample compared to the GE total population a combined χ^2 statistic; 44.37, with a p -value of 0.000013 ($p < 0.001$), showed strong evidence in overall frequency difference between the endometriosis cohort and the GE total population across all endometriosis variants. This suggests that the genetic profile of the endometriosis cohort significantly deviates from the general population, potentially indicating a unique genetic signature associated with endometriosis. Furthermore, the combined χ^2 statistic for the random sample cohort, 12.23, with a p -value of 0.427, showed no significant difference between the random sample cohort and the GE total population across all endometriosis variants. This lack of significant deviation suggests that the random sample's variant frequencies are representative of the GE total population, serving as a control group.

Table 1 shows the variant profiles for the significant variants found within the endometriosis cohort.

LD analysis

LD analysis revealed a linkage between Neandertal-derived *IL-6* variants (rs34880821 and rs2069840), in East Asians ($D' = 1.0$, $r^2 = 0.9662$), suggesting selective retention and potential immune regulation effects (depicted in Fig. 3), Europeans showed LD ($D' = 0.9234$, $r^2 = 0.5794$). The *IL-6* variant rs34880821, located at a Neandertal-derived methylation site, exhibits LD in East Asians ($r^2 = 0.9662$, $D' = 1.0$), while showing LD in Europeans ($r^2 = 0.5794$) and South Asians ($r^2 = 0.8104$) [27]. In African populations, *IL-6* (rs2069840–rs34880821) shows $D' \approx 0.875$, $r^2 \approx 0.482$, while *CNR1* (rs76129761–rs806372) is near-zero, likely due to the absence of Neandertal introgression.

The Denisovan-influenced *CNR1* variant (rs806372) exhibited LD in East Asians ($D' = 0.8004$) but LD in Europeans ($D' = 0.4607$) and South Asians ($D' = 0.1459$), suggesting population-specific evolutionary pressures. African populations had negligible LD for both variant pairs, supporting the hypothesis that these associations arose post-migration due to archaic human introgression.

These findings highlight six regulatory variants significantly enriched in the endometriosis cohort. Notably, the *IL-6* variants rs2069840 and rs34880821 not only co-occurred at a rate well above chance but also exhibit population-specific LD consistent with archaic hominin ancestry and potential immunoregulatory roles. Other significant variants in *CNR1*, *IDO1*, and *KISS1R* were

associated with transcription factor binding sites and are in genomic regions shown to be responsive to EDCs.

Consistent with LD patterns, *CNR1* rs806372 showed a strong EAS-specific PBS (~ 0.49), with much smaller values in AFR and EUR. See supplementary material for more detail.

DISCUSSION

Many sufferers of endometriosis are subject to misdiagnosis and delays due to limited understanding of early-stage risk factors. To obtain a better understanding of genetic factors predisposing to endometriosis development, two literature searches and interrogation of the GE 100,000 genome database were conducted.

Using a highly targeted multilevel approach, five genes have been characterised as potential targets in the developing endometriosis pathway, containing variants or having altered expression levels, with six variants found as statistically significant in our cohort population when compared to GE. This builds on the work conducted by Sapkota, Rahmiloglu, and Zondervan [7–9], who collectively found forty-two SNPs linked to endometriosis.

IDO1

The downstream variant of *IDO1* rs72643906 is found to be rare in the general population, and is significantly higher in the study cohort, indicating the variant may influence endometriosis development through altering *IDO1* expression levels. Specifically, Brooks [28] found that immune system activation increases *IDO1* expression. Furthermore, the UCSC [23] database showed an association with an elevated risk of COVID-19, and therefore, is potentially disadvantageous to the immune system, which is seen to be associated with endometriosis patients [4]. The UCSC database [23] showed rs72643906 to alter the motif of the transcription factor ZIC2, which regulates tissue expression, this potentially modifies the ZIC2 motif to prevent the transcription factor from binding, therefore hypothetically increasing the expression levels of *IDO1* and ZIC2. Furthermore, elevated *IDO1* expression has been found in endometriosis patients [29], showing a potential link between rs72643906 and endometriosis. However, when Bisphenol A (BPA) is introduced through the environment, *IDO1* and ZIC2 expression levels are decreased [30, 31] potentially creating a new dysregulated baseline effect of rs72643906. EDC exposures are considered baseline shifting rather than uniformly opposing variant effects: in carriers, exposures can establish a new dysregulated set-point, BPA lowering *IDO1/ZIC2* despite a variant-predisposed higher baseline and related to possible predictions of exposure dependent penetrance with some additivity.

CNR1

The *CNR1* variants rs76129761 and rs806372 were found to be significant within the study cohort; rs76129761 is deleterious, and rs806372 has a potential splicing site implicating three alternative transcripts within intron 1 of *CNR1*. Therefore, it is possible these variants potentially increase *CNR1* expression levels, leading to endometriosis development, with Allam [32] finding increased *CNR1* levels in endometriosis patients when compared to controls.

Variant rs76129761 was found to be high in the random control group as well as the endometriosis cohort. This may be due to undiagnosed subclinical conditions. Furthermore, the UCSC database [23] found this variant to be high within Northern Sweden, Estonian, and Korean populations and is highly conserved in humans, rhesus monkeys, mice, and dogs. This potentially indicates an advantageous effect of this variant which when exposed to environmental factors may become disadvantageous.

The UCSC database [23] shows rs806372 to be a Denisovan variant, which may be well established in populations where Denisovans and sapiens interact, being randomly selected until

Table 1. Variant profiles for the significant variants found within this cohort (created using UCSC [23]).

Variant	Associated gene.	Variant type/consequence.	Chromosome location.	Potential transcription factor (TF) binding site affected (if applicable).	Number of publications.	Other significant features.	Pollutant interference.	P-value from χ^2 goodness of fit	BH-corrected p-value
rs72643906	<i>IDO1</i>	A > G Downstream variant	Chr8:39779157–39779157	<i>ZIC2</i> (Zic family member 2)	None	Highly conserved in mammals and vertebrates	BPA and TBBPA (tetrabromobisphenol A) can decrease methylation of <i>IDO1</i> 's mRNA and DNA [31], and TCDD can increase <i>IDO1</i> expression levels [49]. Exposure to BPA has been shown to decrease <i>ZIC2</i> expression [30].	0.00175665990877229	0.006
rs933717388	<i>KISS1R</i>	C > G Intron variant	Chr19:912365–912365	<i>ZNF707</i> (Zinc finger protein 707) and <i>ZBTB11</i> (Zinc finger and BTB domain-containing protein 11)	None	Conserved in humans	DEHP exposure influences regulators of HPGA by regulating GnRH (Gonadotropin hormone-releasing hormone) release altering <i>KISS1R</i> expression [42].	0.017530950708949900	0.036
rs34880821	<i>IL-6</i>	G > A Intergenic	Chr7:22775450–22775450	None	3	Conserved in humans Methylation site. Neanderthal variant.	Exposure to TCDD can increase the expression of <i>IL-6</i> [49].	0.006042586245540780	0.024
rs2069840	<i>IL-6</i>	C > G Intron	Chr7:22768572–22768572	None	12	Conserved in rhesus, mouse, and dogs.	Exposure to TCDD can increase the expression of <i>IL-6</i> [49].	0.020059606754631400	0.040
rs76129761	<i>CNR1</i>	CTCT > CT Intron	Chr6:88860156–88860156	<i>ZNF701</i> (Zinc finger protein 701), <i>SOX4</i> (SRY-box transcription factor 4), <i>SOX6</i> (SRY-box transcription factor 6), <i>FOXD3</i> (Forkhead box D3) and <i>SOX11</i> (SRY-box transcription factor 11).	None	Conserved in human, rhesus, mouse, and dogs.	Exposure to DINP and BPA significantly increased the expression of <i>CNR1</i> [34]. Exposure to BPA can upregulate <i>SOX4</i> expression [38], <i>SOX6</i> and <i>SOX11</i> [37]. Exposure to BPA has been shown to decrease <i>FOXD3</i> levels [39].	0.023331593081759400	0.046
rs806372	<i>CNR1</i>	C > G Intron	Chr6:88846563–88846563	<i>SOX12</i> (SRY-box transcription factor 12).	3	Conserved in humans and rhesus. Splicing site; SIB locus ID: NC-000006-1470.3 alternative transcripts within intron 1 (6q15) of <i>CNR1</i> . Denisovan variant	Exposure to DINP and BPA significantly increased the expression of <i>CNR1</i> [34]. DEHP can increase <i>SOX12</i> expression levels [35].	0.003312817523356740	0.018



Fig. 3 LD analysis of variants rs34880821 and rs2069840.

the onset of EDCs in our modern society [33]. The consequence of this may have been exposure to Di-(2-ethylhexyl) phthalate (DEHP) and BPA, which has been found to increase *CNR1* expression [34].

The UCSC database [23] also shows rs806372 to lie on the binding site of SOX12. This may alter the motif, disrupt its binding and lead to increased *CNR1* and decreased SOX12 expression levels. However, environmental DEHP exposure has been shown to elevate both SOX12 and *CNR1* levels, as supported by expression studies [35, 36].

Variant rs76129761 lies on the binding site of five transcription factors; ZNF701, SOX4, SOX6, FOXD3 and SOX11, which may cause increased expression levels of the these and *CNR1* gene. Furthermore, environmental BPA exposure increases SOX4, SOX6, and SOX11 [37, 38] and decreases FOXD3 expression levels [39].

These variants, when exposed to DEHP and BPA may establish dysregulation leading to endometriosis development. EDC exposures are considered baseline shifting rather than uniformly opposing variant effects: in carriers, exposures can establish a new dysregulated set-point DEHP elevating both SOX12 and *CNR1* despite a motif-disrupting allele and related to possible predictions of exposure dependent penetrance with some additivity.

KISS1R

The *KISS1R* intronic variant rs933717388 is a potentially novel variant (no literature or records referencing this variant) and was found to be significantly higher in the study population. The UCSC database [23] found rs933717388 to lie on the binding sites of ZNF707 and ZBTB11. The variant rs933717388 may bind to the sites of transcription factors ZNF707 and ZBTB11 altering their motifs and potentially hindering their binding. Since ZBTB11 is a silencing transcription factor, reduced binding could increase *KISS1R* expression. This upregulation may disrupt reproductive function and promote endometriotic cell metastasis [40], Blasco [41] supports this, reporting elevated *KISS1R* levels in granulosa cells of endometriosis patients. Both *KISS1R* and transcription factor levels are potentially increased in the presence of rs933717388, and DEHP exposure further raises *KISS1R* expression [42], though its effect on ZNF707 and ZBTB11 remains unstudied.

IL-6

Intronic variants rs2069850 and rs34880821 of the *IL-6* gene were found to be significant within the study cohort. These variants

have been found to colocalise in the endometriosis cohort, which suggests a potential biological interaction between these variants that are potentially relevant to endometriosis. UCSC [23] showed rs34880821 to be a Neandertal methylation site. Therefore, the reference allele was methylated in Neandertals and Denisovans, but this variant abolishes the methylation site, so rs34880821 may exacerbate endometriosis development due to the disease being associated with dysregulation of the immune and inflammatory system by aberrant silencing of the area and continuous expression. This probability of heightened *IL-6* expression in endometriosis patients is further validated in research by Li [43], finding an increase in *IL-6* expression levels when compared to controls.

The intersection of ancient genetic variants, epigenetic regulation, and modern environmental pollutants in endometriosis susceptibility

Our findings highlight how ancient genetic variants inherited from Neandertals and Denisovans, epigenetic regulation, and modern industrial pollutants may potentially converge to shape population-specific risks for endometriosis. The LD analysis of *IL-6* and *CNR1* regulatory variants suggests that evolutionary pressures involving immune regulation and inflammatory responses, when combined with modern environmental exposures, may amplify disease susceptibility [27, 44].

Neandertal introgression has significantly influenced *IL-6* regulation, particularly in East Asian populations. This suggests that Neandertal introgression may have played a role in shaping *IL-6* regulatory pathways, particularly in East Asians, where it remains strongly linked. Given that EAS also have the highest reported prevalence of endometriosis (~15.4%), it is likely that these genetic variants contribute to heightened inflammatory responses, immune dysregulation, and fibrotic lesion formation [45].

The functionality for the Neandertal-derived methylation at rs34880821 may indicate another subtle contributor to endometriosis risk. Methylation usually functions as a gene-silencing mechanism, regulating cytokine levels to prevent excessive inflammation [46]. If methylation is lost, *IL-6* expression could become hyperactive, leading to chronic immune activation, sustaining peritoneal inflammation, fibrosis, and deep-infiltrating endometriosis, as *IL-6* is known to drive fibrotic remodelling in reproductive tissues. It also may account for heightened neuroinflammation, potentially explaining increased pain sensitivity in East Asian endometriosis patients [45].

Similarly, the Denisovan-derived rs806372 variant in *CNR1*, which is involved in immune modulation and pain perception, exhibits LD in East Asians [26]. Since *CNR1* plays a key role in pain signalling and inflammatory responses, this Denisovan-derived variant may enhance pain sensitivity in individuals with endometriosis, particularly in EAS [47].

While these ancestral variants may have once provided immune advantages in prehistoric environments, modern environmental factors may be reversing these evolutionary benefits, transforming once-adaptive immune responses into drivers of chronic disease [45].

Exposure to EDCs, such as BPA, phthalates, and dioxins, has been shown to demethylate immune regulatory genes, including *IL-6*, leading to excessive cytokine production [48]. If Neandertal-derived *IL-6* regulatory variants are already prone to overactivation, additional EDC-induced demethylation could further escalate inflammatory signalling, worsening lesion development. This suggests a gene-environment interaction, where modern industrial chemicals exacerbate genetic predispositions inherited from archaic human ancestors.

In contrast to EAS and EUR, AFR exhibit significantly lower LD for these *IL-6* and *CNR1* variants ($D' < 0.02$, $r^2 < 0.001$), suggesting these genetic associations arose post-migration due to Neandertal and Denisovan introgression [27]. There is limited available data on the prevalence of endometriosis in AFR, and estimates may vary due to underdiagnosis.

PBS profiles indicate that the *IL-6* regulatory pair is most differentiated in EUR and the *CNR1* signal is strongest in EAS, consistent with the directionality of the LDpop-derived allele frequency differences. Although PBS alone does not prove selection there is concordance with functional hypotheses that warrants further study.

This study supports a novel perspective for endometriosis susceptibility, in which ancestral genetic variants interact with modern environmental pollutants to modulate immune regulation, chronic inflammation, and pain perception across populations.

Limitations and future directions

While this study provides novel insights into the genetic and environmental factors influencing endometriosis susceptibility, several limitations must be acknowledged. We reference archaic ancestry as context (haplotypic/methylation signatures) rather than proof of inheritance. Future work will include outgroup allele-state checks and selection statistics to test evolutionary hypotheses directly. The sample size was limited, particularly in younger individuals under twenty-nine years old. Additionally, family-based cascade genetic testing was not included, preventing the evaluation of heritability and potential familial aggregation of risk variants. Another limitation is the reliance on medical records to infer endometriosis staging, which introduces the potential for misclassification bias. Furthermore, while this study identified significant associations between regulatory variants and endometriosis risk, functional validation through in vitro and vivo models is required to confirm their biological relevance.

Further research should focus on expanding the study cohort to include a larger and more diverse population to improve the statistical power of genetic associations and assess whether these findings are consistent across different ethnic backgrounds. Family-based studies incorporating cascade genetic testing could help clarify inheritance patterns and the potential contribution of additional rare variants to endometriosis risk. Given the discovery phase design, findings are hypothesis-generating. Replication in large, ancestry-stratified cohorts (e.g., UK Biobank/other databases) will be pursued in a subsequent protocol and is beyond the scope of this pilot. Further functional studies are necessary to evaluate how the identified regulatory variants influence gene expression and immune signalling pathways, particularly in

response to EDCs. Integrating environmental exposure data with genomic analysis would provide a more comprehensive understanding of how genetic and environmental factors interact in the development and progression of endometriosis. Potentially enabling computational modelling of gene regulatory networks under varying EDC exposures to predict combined outcomes and guide specific functional validation experiments. These future directions will be critical for translating genetic discoveries into practical applications for early diagnosis and personalised treatment strategies.

CONCLUSION

This study provides a novel perspective on the genetic and environmental interplay driving endometriosis susceptibility, highlighting the influence of ancient regulatory variants and modern industrial pollutants. This research identified statistically significant regulatory variants in *IL-6*, *CNR1*, *IDO1*, and *KISS1R* that may contribute to endometriosis risk through interactions with EDCs. These findings aim to bridge the gap between genetic predisposition, and environmental exposures, shedding light on how Neandertal and Denisovan introgressed variants in immune and pain-regulatory genes may influence disease susceptibility in modern populations.

This is the first study to systematically explore how ancient genetic signatures, epigenetic regulation, and contemporary environmental pollutants intersect in the pathophysiology of endometriosis. These results lay the first steps for future precision medicine approaches, where genetic screening combined with environmental risk assessment may improve early detection and personalised intervention strategies. Moving forward, functional validation of these regulatory variants and their interaction with endocrine disruptors will be crucial in understanding their mechanistic role in endometriosis onset and progression. These findings enhance our understanding of endometriosis as a multifactorial disease but also provide a framework for future research integrating evolutionary genetics, environmental health, and reproductive medicine.

DATA AVAILABILITY

Research on the de-identified patient data used in this publication can be carried out in the Genomics England Research Environment subject to a collaborative agreement that adheres to patient led governance. All interested readers will be able to access the data in the same manner that the authors accessed the data. For more information about accessing the data, interested readers may contact research-network@genomicsengland.co.uk or access the relevant information on the Genomics England website: <https://www.genomicsengland.co.uk/research>.

REFERENCES

1. Kanellopoulos D, Karagianni D, Pergialiotis V, Patsouras G, Patsouras K, Nikiteas N, et al. The interplay between endometriosis and fertility in rats: a systematic review. *J Med Life*. 2022;15:742.
2. Agarwal SK, Chapron C, Giudice LC, Laufer MR, Leyland N, Missmer SA, et al. Clinical diagnosis of endometriosis: a call to action. *Am J Obstet Gynecol*. 2019;220:170–71.
3. Nnoaham KE, Hummelshoj L, Webster P, d'Hooghe T, de Cicco Nardone F, de Cicco Nardone C, et al. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. *Fertil Steril*. 2011;96:366–73.
4. García-Gómez E, Vázquez-Martínez ER, Reyes-Mayoral C, Cruz-Orozco OP, Camacho-Arroyo I, Cerbón M. Regulation of Inflammation Pathways and Inflammasome by Sex Steroid Hormones in Endometriosis. *Front Endocrinol*. 2020;10:935.
5. Saha R, Pettersson HJ, Svedberg P, Olovsson M, Bergqvist A, Marions L, et al. Heritability of endometriosis. *Fertil Steril*. 2015;104:947–52.
6. La Merrill MA, Vandenberg LN, Smith MT, Goodson W, Browne P, Patisaul HB, et al. Consensus on the key characteristics of endocrine-disrupting chemicals as a basis for hazard identification. *Nat Rev Endocrinol*. 2020;16:45–57.

7. Sapkota Y, Steinhorsdottir V, Morris AP, Fassbender A, Rahmioglu N, De Vivo I, et al. Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat Commun*. 2017;8:15539.
8. Rahmioglu N, Mortlock S, Ghiasi M, Möller PL, Stefansdottir L, Galarneau G, et al. The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions. *Nat Genet*. 2023;55.
9. Zondervan KT, Becker CM, Koga K, Missmer SA, Taylor RN, Viganò P. Endometriosis. *Nat Rev Dis Primers*. 2018 Dec 1 [cited 2024 Jul 17];4. Available from: <https://pubmed.ncbi.nlm.nih.gov/30026507/>.
10. The National Genomic Research Library v5.1. 2020 [cited 2025 May 10]; Available from: www.genomicsengland.co.uk.
11. 100,000 Genomes Project | Genomics England [Internet]. [cited 2024 Jul 17]. Available from: <https://www.genomicsengland.co.uk/initiatives/100000-genomes-project>.
12. You HH, Song G. Review of endocrine disruptors on male and female reproductive systems. Vol. 244, *Comparative Biochemistry and Physiology Part - C: Toxicology and Pharmacology*. 2021.
13. Ensembl genome browser 112 [Internet]. [cited 2024 Jul 17]. Available from: <https://www.ensembl.org/index.html>.
14. Home - GraphPad [Internet]. [cited 2024 Jul 17]. Available from: <https://www.graphpad.com/>.
15. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. *Front Genet* [Internet]. 2020;11:157. <https://doi.org/10.3389/fgene.2020.00157>.
16. Machiela MJ, Chanock SJ. LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. *Bioinformatics* [Internet]. 2018;34:887–9. <https://doi.org/10.1093/bioinformatics/btx561>.
17. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31:3555–7. <https://doi.org/10.1093/bioinformatics/btv402>.
18. 'c.1018delG' - ClinVar - NCBI [Internet]. [cited 2024 Jul 17]. Available from: <https://www.ncbi.nlm.nih.gov/clinvar/?term=c.1018delG>.
19. 672[genid] - ClinVar - NCBI [Internet]. [cited 2024 Jul 17]. Available from: [https://www.ncbi.nlm.nih.gov/clinvar/?term=672\[genid\]](https://www.ncbi.nlm.nih.gov/clinvar/?term=672[genid]).
20. 17[chr] AND 43000000:44000000[chrpos37] - ClinVar - NCBI [Internet]. [cited 2024 Jul 17]. Available from: <https://www.ncbi.nlm.nih.gov/clinvar/?term=17%5Bchr%5d+AND+43000000:44000000%5Bchrpos37>.
21. Sherry ST, Ward M, Sirotkin K. dbSNP—Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation. *Genome Res* [Internet]. 1999;9:677–9. <https://genome.cshlp.org/content/9/8/677>.
22. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Database issue Published online* [Internet]. 2023 [cited 2025 May 10];51. Available from: <https://doi.org/10.1093/nar/gkac1000>.
23. Nassar LR, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, et al. The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res* [Internet]. 2023;51:D1188–95. <https://pubmed.ncbi.nlm.nih.gov/36420891>.
24. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2022;50:687–92.
25. SNPedia [Internet]. [cited 2024 Jul 17]. Available from: <https://www.snpedia.com/>.
26. National Library of Medicine - National Institutes of Health.
27. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*. 2014;507:354–7. <https://www.nature.com/articles/nature12961>.
28. Brooks AK, Lawson MA, Smith RA, Janda TM, Kelley KW, McCusker RH. Interactions between inflammatory mediators and corticosteroids regulate transcription of genes within the Kynurenine Pathway in the mouse hippocampus. *J Neuroinflammation*. 2016;13:98.
29. Mei J, Li MQ, Ding D, Li DJ, Jin LP, Hu WG, et al. Indoleamine 2,3-dioxygenase-1 (IDO1) enhances survival and invasiveness of endometrial stromal cells via the activation of JNK signaling pathway. *Int J Clin Exp Pathol*. 2013 [cited 2024 Jul 17];6:431.
30. Brown JS. Effects of bisphenol-A and other endocrine disruptors compared with abnormalities of schizophrenia: An endocrine-disruption theory of schizophrenia. *Schizophrenia Bull*. 2009;35:256–78.
31. Merrill AK, Sobolewski M, Susiarjo M. Exposure to endocrine disrupting chemicals impacts immunological and metabolic status of women during pregnancy. *Mol Cell Endocrinol*. 2023;577:112031.
32. Allam S, Paris E, Lazzano I, Bitterman P, Basu S, O'Donnell J, et al. Detection of cannabinoid receptor expression by endometriotic lesions in women with endometriosis as an alternative to opioid-based pain medication. *J Immunol Res*. 2022;2022:4323259.
33. Zhang X, Witt KE, Bañuelos MM, Ko A, Yuan K, Xu S, et al. The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. *Proc Natl Acad Sci USA*. 2021;118:e2020803118.
34. Forner-Piquer I, Santangeli S, Maradonna F, Verde R, Piscitelli F, di Marzo V, et al. Role of Bisphenol A on the Endocannabinoid System at central and peripheral levels: Effects on adult female zebrafish. *Chemosphere*. 2018;205:118–25.
35. Chou CK, Huang HW, Yang CF, Dahms HU, Liang SS, Wang TN, et al. Reduced camptothecin sensitivity of estrogen receptor-positive human breast cancer cells following exposure to di(2-ethylhexyl)phthalate (DEHP) is associated with DNA methylation changes. *Environ Toxicol*. 2019;34:401–14.
36. Ernst J, Grabiec U, Falk K, Dehghani F, Schaedlich K. The endocrine disruptor DEHP and the ECS: Analysis of a possible crosstalk. *Endocr Connect*. 2020;9:101–10.
37. Lichtensteiger W, Bassetti-Gaille C, Rehrauer H, Georgijevic JK, Tresguerres JAF, Schlumpf M. Converging effects of three different endocrine disruptors on sox and pou gene expression in developing rat hippocampus: possible role of microRNA in sex differences. *Front Genet*. 2021;12:718796.
38. Kang SY, Song JY, Cho HH. Gene expression analysis of uterine smooth muscle cells exposed to bisphenol A. *Toxicol Environ Health Sci*. 2014;6.
39. Baba K, Okada K, Kinoshita T, Imaoka S. Bisphenol A disrupts notch signaling by inhibiting gamma-secretase activity and causes eye dysplasia of *Xenopus laevis*. *Toxicological Sciences*. 2009;108.
40. Xie Q, Kang Y, Zhang C, Xie Y, Wang C, Liu J, et al. The role of kisspeptin in the control of the hypothalamic-pituitary-gonadal axis and reproduction. *Front Endocrinol*. 2022;13:925206.
41. Blasco V, Pinto FM, Fernández-Atucha A, González-Ravina C, Fernández-Sánchez M, Cadenas L. Female infertility is associated with an altered expression of the neurokinin B/neurokinin B receptor and kisspeptin/kisspeptin receptor systems in ovarian granulosa and cumulus cells. *Fertil Steril*. 2020;114:869–78.
42. Yu Z, Wang F, Han J, Lu R, Li Q, Cai L, et al. Opposite effects of high- and low-dose di-(2-ethylhexyl) phthalate (DEHP) exposure on puberty onset, oestrous cycle regularity and hypothalamic kisspeptin expression in female rats. *Reprod Fertil Dev*. 2020;32:610–8.
43. Li C, Zhao HL, Li YJ, Zhang YY, Liu HY, Feng FZ, et al. The expression and significance of leukemia inhibitory factor, interleukin-6 and vascular endothelial growth factor in Chinese patients with endometriosis. *Arch Gynecol Obstet*. 2021;304:163–70. <https://link.springer.com/article/10.1007/s00404-021-05980-5>.
44. Harris K, Nielsen R. The Genetic Cost of Neanderthal Introgression. *Genetics* [Internet]. 2016;203:881–91. <https://doi.org/10.1534/genetics.116.186890>.
45. Velarde MC, Bucu MEM, Habana MAE. Endometriosis as a highly relevant yet neglected gynecologic condition in Asian women. *Endocr Connect*. 2023 Nov 1 [cited 2025 May 10];12. Available from: <https://ec.bioscientifica.com/view/journals/ec/12/11/EC-23-0169.xml>.
46. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* 2020 587:7835 [cited 2025 May 10];587:610–2. Available from: <https://www.nature.com/articles/s41586-020-2818-3>.
47. Yen CF, Kim MR, Lee CL. Epidemiologic factors associated with endometriosis in East Asia. Vol. 8, *Gynecology and minimally invasive therapy*. 2019;8:4–11.
48. Liu Z, Lu Y, Zhong K, Wang C, Xu X. The associations between endocrine disrupting chemicals and markers of inflammation and immune responses: A systematic review and meta-analysis. *Ecotoxicol Environ Saf*. 2022;234:113382.
49. Matta K, Koual M, Ploteau S, Coumoul X, Audouze K, Le Bizet B, et al. Associations between exposure to organochlorine chemicals and endometriosis: a systematic review of experimental studies and integration of epidemiological evidence. *Environ Health Perspect*. 2021;129:076003.

ACKNOWLEDGEMENTS

Posthumous acknowledgement. This work is dedicated to our beloved colleague Dr Elpida Fragkouli for her consistent support, supervision and contribution to the study design and data interpretation. This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure.

AUTHOR CONTRIBUTIONS

AW contributed to the study design, collected the data and analysed the data and drafted the manuscript. DA contributed to the study design, provided supervision of

the study and contributed to the article drafting and editing. DW contributed to the study design, analysed the data and editing. JW contributed to data gathering. AM leads the GE project that this study is a part of, designed the study, data analysis and interpretation, provided supervision of the study and manuscript drafting and editing.

FUNDING

Bournemouth University funded the writing of this manuscript granted to the first author of this manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICAL APPROVAL

This study received ethical approval from Bournemouth University's Institutional Research Ethics Panel (ID: 45978). Genomic data were obtained under Project ID 645 through the Genomics England 100,000 Genomes Project. All participants provided informed consent for the use of their data in secondary research.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41431-025-01977-9>.

Correspondence and requests for materials should be addressed to Anna Mantzouratou.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025