Genetic variation in life-history traits is not correlated with estrogen-receptor positive breast cancer risk

Euan A Young, E Yagmur Erten, Erik Postma, Virpi Lummaa, Bert van der Vegt, Geertruida H de Bock, and Hannah L Dugdale

Author names, ORCID ID

Euan A. Young[1]*, 0000-0001-9370-9681

E Yagmur Erten[1], 0000-0001-9736-0405

Erik Postma[2], 0000-0003-0856-1294

Virpi Lummaa[3],

Bert van der Vegt[4], 0000-0002-2613-1506

Geertruida H de Bock[5], 0000-0003-3104-4471

Hannah L. Dugdale[1]*, 0000-0001-8769-0099

1 - Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

2 - Centre for Ecology and Conservation, University of Exeter, Penryn, United Kingdom

3 - Department of Biology, University of Turku, Turku, Finland

4 – Department of Pathology & Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

5- Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

* Corresponding author emails: e.a.young@rug.nl, h.l.dugdale@rug.nl

ABSTRACT

Estrogen-receptor positive breast cancer (ER+ BC) is one of the most prevalent cancers, but the evolutionary processes shaping genetic variation in ER+ BC risk are poorly understood. Here, we explore evidence for antagonistic pleiotropy between ER+ BC risk and human life-history traits, testing whether increased genetic ER+ BC risk is indicative of life-history trade-offs, and is associated with faster maturation, earlier reproduction, and/or increased reproductive success. We estimate genome-wide genetic correlations between ER+ BC risk and three life-history traits (age at menarche, age at first birth, and the number of children born) for Lifelines biobank women using genomic restricted maximum likelihood analyses. We complement these by – for the first time – performing linkage disequilibrium score regression on European family-based genome-wide association study data. Regardless of the data or method used, genetic correlations were low and not statistically different from zero. Thus, we find no evidence for antagonistic pleiotropy between ER+ BC risk and life-history traits, suggesting that these germline mutations have instead accumulated due to weak selection at older ages. Our quantitative genetic analyses suggest that previous results showing antagonistic pleiotropy for individual genetic variants do not replicate when accounting for these traits' polygenic nature and that observed phenotypic trends do not have a genetic underpinning. While we find no evidence for life-history trade-offs shaping genetic risk for ER+ BC, our rigorous approach may provide a template for future studies to test the role of life-history trade-offs in shaping genetic risk for other diseases.

KEYWORDS: mutation accumulation, life-history trade-offs, evolutionary medicine, quantitative genetics, SNP heritability, antagonistic pleiotropy

INTRODUCTION

Breast cancer is the fourth most deadly cancer worldwide (GBD, 2021) and accounts for the greatest proportion of new cancer cases (Sung et al., 2021). Estrogen receptor-positive breast cancer (ER+ BC) is the most common subtype, accounting for ~80% of all breast cancer cases and is most common after age ~50 (Yasui & Potter, 1999). While lifestyle can influence an individual's risk of developing ER+ BC (Colditz et al., 2004), common germline genetic variation explains ~14% of the variation in a woman's risk (Zhang et al., 2020) and over 200 genome-wide significant risk variants have now been identified (Michailidou et al., 2013, 2017; Zhang et al., 2020). However, we have a poor understanding of why genetic variants promoting a life-threatening disease persist in contemporary populations. Examining how ER+ BC risk variants are associated with human life-history traits would give insight into the evolutionary processes giving rise to these variants, their genetic architecture, and help answer whether these risk variants will become more or less prevalent in future generations.

Organisms have been under selective pressure to evolve cancer suppression mechanisms since the dawn of multicellularity over half a billion years ago (Merlo et al., 2006; Nunney, 2013), yet cancer remains common across the tree of life (Aktipis, Boddy, et al., 2015). Two non-mutually exclusive evolutionary hypotheses can help explain why germline risk variants persist in human populations. First, they can accumulate due to weak selection at older ages, where there are fewer individuals for selection to act upon and most individuals have already reproduced (*mutation accumulation theory*, Figure 1; Charlesworth, 1980, 2001; Haldane, 1941; Medawar, 1952). Second, risk variants may be selected for because they are associated with increased fitness due to their beneficial effects at younger ages, as a part of life-history trade-offs (*antagonistic pleiotropy theory*, Figure 1; Williams, 1957). Indeed, life-history theory suggests that cancer suppression mechanisms are costly because they may trade-off against other functions, such as growth and reproduction, as organisms evolve to find an optimal balance between competing demands under limited resources (Jacqueline et al., 2017; Roff, 1992; Stearns, 1977, 1992). Thus, genes promoting a faster life-history (earlier maturation and reproduction) could be selected for if they result in higher reproductive success, even if they increase the risk of cancer in later life (Dujon et al., 2022; Jacqueline et al., 2017; Promislow & Harvey, 1990). Such antagonistically pleiotropic genes have been found across an array of human diseases (e.g., Rodríguez et al., 2017; Trumble et al., 2023, reviewed in Byars & Voskarides, 2020), including cancer (Crespi & Summers, 2006). However, although there is evidence for a reproduction-cancer suppression trade-off amongst species (e.g., in birds, Kapsetaki et al., 2024), the role of antagonistic pleiotropy in shaping genetic variation in ER+ BC risk in humans is not well understood.

Human life-history traits have been widely demonstrated to associate ER+ BC risk in phenotypic studies with earlier ages at menarche, later ages at first birth, and lower numbers of children associated with

increased ER+ BC risk (Figure 1, Aktipis, Ellis, et al., 2015; Kelsey & Bernstein, 1996; Momenimovahed & Salehiniya, 2019). This may be due to these trends resulting in women spending less time in amenorrhea, experiencing more menstrual cycles, and having higher lifetime exposure to estrogen (Eaton et al., 1994; Short, 1974; Strassmann, 1999), which has proliferative effects on mammary cells and thereby increases the risk of tumor development (Dall et al., 2018). ER+ BC risk being associated with fewer children furthermore suggests that ER+ BC risk variants are currently under negative selection in contemporary populations (Figure 1). However, the influence of environmental factors on both ER+ BC risk and human life-history traits (such as lifestyle factors and socioeconomic and cultural backgrounds, Barban et al., 2016; Colditz et al., 2004) means that phenotypic trends cannot predict how human life-history traits might influence ER+ BC risk at the genetic level, and it is only by studying genetic associations that we can establish if these phenotypic trends translate into evolutionary change (Haldane, 1954; Lande & Arnold, 1983; Robertson, 1966).

Genetic studies of how human life-history traits influence ER+ BC risk remain scarce. One genetic study using genealogy data from Utah found some evidence of antagonistic pleiotropy maintaining mutations in BRCA1/2 genes, with women carrying these breast cancer risk variants also having more children (Smith et al., 2012). Additionally, for the 375 most important loci for age at menarche, variants lowered age at menarche while increasing subsequent ER+ BC risk (Day et al., 2017). However, studies which focus on specific variants fail to capture the full scope of genetic variation in both ER+ BC risk and human life-history traits. For example, the 375 loci used in this study amounted to only 25% of the total genetic variation in age at menarche (Day et al., 2017), with the hundreds of genome-wide significant variants for ER+ BC risk explaining less than 50% of the total genetic variation (Zhang et al., 2020). Studies estimating genome-wide genetic correlations between ER+ BC risk and human life-history traits are limited to two studies estimating genetic correlations between ER+ BC risk and age at menarche (Jiang et al., 2019; Prince et al., 2024), one of which also estimated the genetic correlations between other life-history traits, including age at first birth and number of children (Prince et al., 2024). These studies revealed negligible genetic correlations between all life-history traits and ER+ BC risk, except for a small (r_G=-0.048) and statistically non-significant ($p = 0.062$) genetic correlation between number of children and ER+ BC risk, suggesting weak selection against ER+ BC risk. However, these studies did not account for confounding factors such as population stratification, which can bias genetic correlations (Berg et al., 2019; Border et al., 2022), particularly those involving human life-history traits (Howe et al., 2022; Tan et al., 2024). Thus, it is still poorly understood how the frequency of genetic variants underlying an important disease will change in future generations, and what adaptive processes gave rise to their prevalence.

To resolve this, we first estimate the importance of genome-wide genetic variation in ER+ BC risk and three life-history traits (age at menarche, age at first birth, and the number of children born), and then examine how the genetic variation in life-history traits is associated with genetic variation in ER+ BC risk. Our predictions under different hypotheses are shown in Figure 1. As the number of children is most closely related to fitness, we take significant genetic correlations between ER+ BC risk and the number of children as the strongest evidence for selection on genome-wide ER+ BC risk. We consider positive genetic correlations between the number of children and ER+ BC risk as indicative of antagonistic pleiotropy due to a reproduction-survival trade-off, with negative correlations between ER+ BC risk and both age at menarche and first birth suggesting that higher ER+ BC risk is associated with a fast-life history strategy.

To estimate these genetic correlations, we first perform a genomic restricted maximum likelihood (GREML) analysis (Yang et al., 2010) using the Dutch biobank Lifelines (Scholtens et al., 2015) linked to national pathology records to estimate heritabilities and genetic correlations. To independently replicate our results, while controlling for potential biases, we perform – for the first time for these traits – linkage disequilibrium (LD) score regressions of family-based GWAS data, which exploits the random Mendelian segregation among siblings to remove environmental confounding from genetic correlations (Kong et al., 2018; Young et al., 2022; Young, 2024).

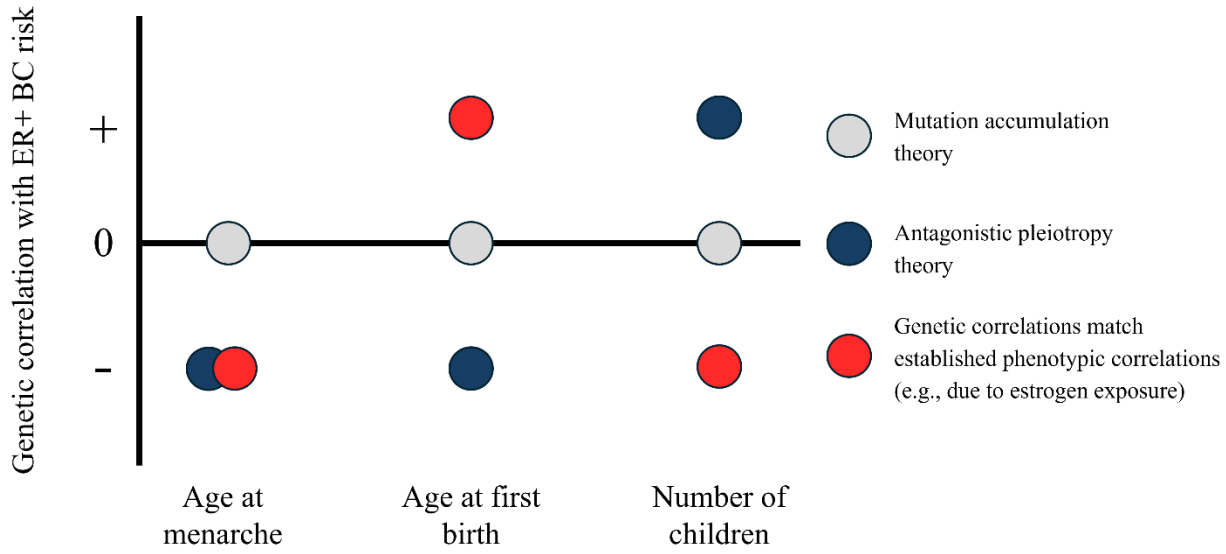

Figure 1: Predicted direction of genetic correlations between ER+ BC risk and life-history traits (age at menarche, age at first birth, and number of children) under different hypotheses. Circles below the x-axis indicate negative genetic correlations, i.e. genetic predispositions for lower/earlier values for the life-history trait are correlated with increased ER+ BC risk. Circles above the x-axis indicate positive genetic

correlations, i.e. genetic predispositions for higher/later values for the life-history trait are correlated with increased ER+ BC risk. Circles on the x axis line show no correlation. Circle colors indicate the expected direction of genetic correlations under different hypotheses: grey, *mutation accumulation theory*, genetic predispositions to ER+ BC risk are not associated with life-history trade-offs; blue, *antagonistic pleiotropy theory*, genetic predispositions to ER+ BC are indicative of a reproduction-survival trade-off and/or faster life history; and red, *genetic correlations match established phenotypic correlations (e.g. due to estrogen exposure)*, genetic correlations match the existing evidence from phenotypic studies of how life-history traits shape ER+ BC risk phenotypically, potentially due to their influence on a woman's lifetime estrogen exposure.

RESULTS

GREML analyses on Lifelines data

We first performed GREML analyses using the Global screening array genomic data available on women from the Lifelines biobank. Univariate analyses revealed statistically significant heritability (±SE) for age at menarche (43.1±5.8%) and age at first birth (24.2±7.4%), suggesting that genetic variance plays a significant role in shaping these traits, but the heritability of the number of children was most likely low (12.8±9.8%, p = 0.337, Table 1a). This was also true for ER+ BC risk, where the heritability was estimated as 11.8% (±5.2%, p = 0.190, Table 1a). These heritability estimates were largely replicated using other genomic data available on Lifelines (the Affymetrix and CytoSNP genomic array data, Table S1), but the heritability estimate for ER+ BC risk became statistically significant when using the Affymetrix array data (p = 0.04, Table S1), probably because these contained a slightly higher number of ER+ BC cases. Furthermore, the heritability estimate for age at menarche was considerably lower (but remained statistically significant) when using the CytoSNP data (17.0% ±5.2%, p < 0.001, Table S1), most likely because of the lower genomic coverage of the CytoSNP data (Table S2).

Table 1: Heritability estimates (and standard error, SE) of traits from **a)** GREML analyses using Lifelines genotype data and **b)** LD score regression on GWAS data. P-values from LRT tests (for GREML) and z-scores (for LD score regressions) and results are presented in bold when significant (< 0.05). Results from Lifelines Global screening array data are shown. Results from analyses of other Lifelines genotype data are shown in Table S1.

	a) GREML on Lifelines data		b) LD score regression on GWAS data	
Trait	Heritability±SE	P-value	Heritability±SE	P-value

ER+ BC risk	11.8±5.2%	0.190	**10.7±0.9%**	**< 0.001**
Age at menarche	**43.1±5.8%**	**< 0.001**	**14.1±5.9%**	**0.008**
Age at first reproduction	**24.2±7.4%**	**< 0.001**	1.6±1.9%	0.202
Number of children	12.8±9.8%	0.337	**8.6±1.7%**	**< 0.001**

Despite demonstrating the importance of genetic variation in shaping these traits, genetic correlations between ER+ BC risk and life-history traits were low and statistically non-significant in the Lifelines data (Figure 2a, Table S3). For the Global screening array data these were 0.05±0.17 for age at menarche, 0.10±0.24 for age at first birth, and 0.10±0.35 for number of children ($p > 0.016$, Fig 2a, Table S2) and these findings were replicated when using both the Affymetrix and CytoSNP data (Table S3).

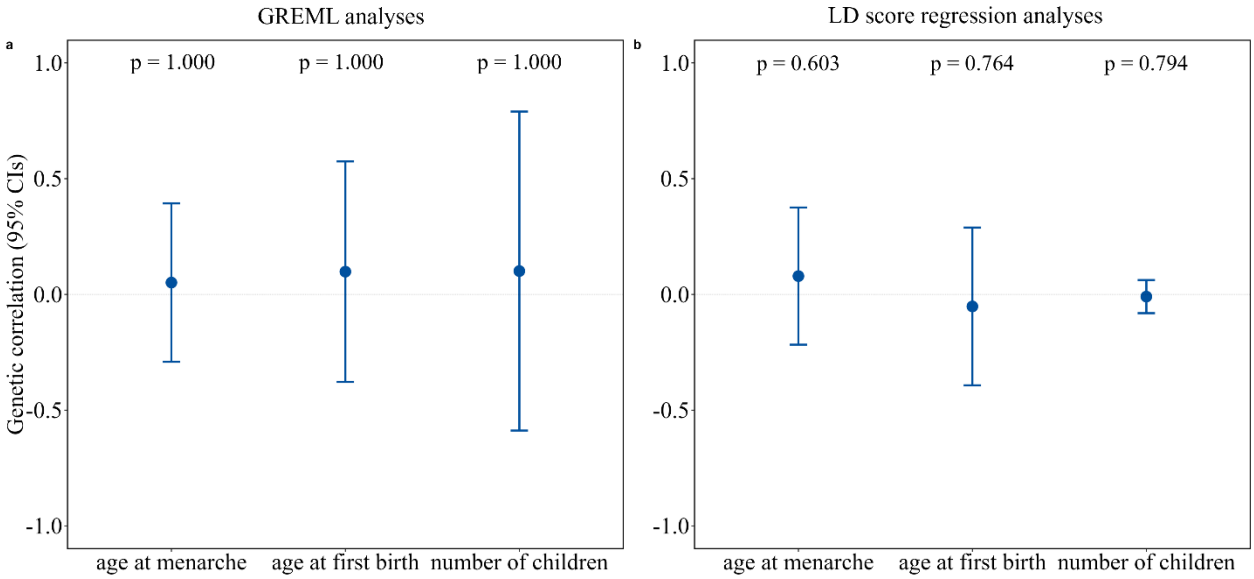

Figure 2: Genetic correlations between ER+ BC risk and life-history traits from: (**a**) a genomic relatedness matrix (GREML) analyses using Lifelines biobank data and (**b**) a LD score regression analyses using family-based GWAS data. Circles above the dotted line indicate positive genetic correlations and circles below indicate negative genetic correlations. Error bars show 95% confidence intervals approximated from SEs. P-values testing a non-zero genetic correlation are shown above each trait.

LD score regressions using GWAS data

GREML analyses can be biased due to gene-environment correlations. To account for these, we repeated our analyses using LD score regression using family-based GWAS data. For this analysis we used European GWAS data from the largest available study on ER+ BC (Michailidou et al., 2017) and the most recent family-based GWAS data study including life-history traits (Tan et al., 2024). Albeit estimated with more precision, our heritability estimates using direct genetic effects from family-based GWAS analyses were largely consistent with our results from the Lifelines biobank for ER+ BC risk (10.7±0.9%) and the number of children (8.6%±1.7%), further suggesting that genetic variation plays a small but non-zero role in shaping among individual variation in these traits (Table 1). However, contrary to results from the Lifelines biobank, the heritability of age at menarche was lower (14.1±5.9% versus 43.1±5.8%), although it remained statistically significant, whereas age at first birth was not shaped by genetic variation when using GWAS data (1.6%±1.9, Table 1). Overall, this suggests that the estimates of the heritability of age at first birth using GREML analyses may have been inflated through factors such as unaccounted population stratification.

Despite these differences, using family-based GWAS summary statistics replicated the finding of no genetic correlations between life-history traits and ER+ BC risk. Our estimates were more precise and smaller than those from the Lifelines biobank analysis, while remaining statistically non-significant (Figure 2b, Table S4). Specifically, we estimated the genetic correlation between ER+ BC risk and our life-history traits as: age at menarche, 0.079 (±0.15, $p = 0.602$); age at first birth, -0.052 (±0.178, $p = 0.764$); and the number of children ever born, -0.010 (±0.037, $p = 0.795$). Taking the 95% confidence intervals, these estimates allowed us to rule out genetic correlations between life-history traits and ER+ BC risk >0.17 and <-0.39 for age at menarche, >0.37 and <-0.22 for age at first birth above, and >0.06 or <-0.08 for the number of children.

DISCUSSION

Our study is the first comprehensive examination of the role of antagonistic pleiotropy and life-history trade-offs in shaping genome-wide genetic variation in risk for one of the most commonly diagnosed cancers, estrogen-receptor positive breast cancer, ER+ BC. Our results provide no support for antagonistic pleiotropy between ER+ BC risk and human life-history traits, suggesting that life-history trade-offs play little role in shaping genome-wide ER+ BC risk, with genetic predispositions to earlier maturation, earlier reproduction and higher reproductive output not also being associated with higher cancer risk. We find no evidence of genome-wide selection on ER+ BC risk variants, suggesting these variants have accumulated due to weak selection at older ages. These findings were replicated, regardless of the analytical approach used, and despite using high-precision genomic, pathology, and the largest available GWAS data. These are the first

such findings using family-based GWAS data and also replicate previous studies using other GWAS data (Jiang et al., 2019; Prince et al., 2024).

Genetic variants increasing the risk of ER+ BC may have been a result of life-history trade-offs for increased reproduction (Smith et al., 2012) or faster life-history strategies (Day et al., 2017). Our quantitative genetic analyses, fully accounting for the polygenic nature of both ER+ BC risk and human life-history traits, suggest that these examples of antagonistic pleiotropy are the exception rather than the rule. For example, while some variants may promote both higher reproductive success and reduced cancer suppression (e.g., BRCA1/2 variants, suggested by Smith et al., 2012), our results suggest that the vast majority of variants show little association with life-history traits resulting in overall no genetic correlation. Alternatively, our results could indicate the complex nature of ER+ BC: genetic variants supporting the antagonistic pleiotropy theory were balanced out by other variants having opposite effects, but this is only speculation. Future studies could examine individual variants in more detail but given we can rule out all but the smallest selection for or against ER+ BC risk variants, our results are most in line with the mutation accumulation theory of ageing. This is despite demonstrating that contemporary European populations retain the ability to evolve through natural selection (Fisher, 1930; Haldane, 1954, Byars et al, 2010). In fact, our heritability estimate of ~8.6% for the number of children is higher than the average estimates of heritability of fitness among wild bird and mammal species (3%, Bonnet et al., 2022). Thus while previous phenotypic studies suggest that ER+ BC risk variants are under negative selection (Aktipis, Ellis, et al., 2015), our studies show that these phenotypic trends do not translate to the genetic level and that these variants will not significantly increase or decrease in prevalence in future generations due to selection. However, we also cannot rule out the effects of historical selection on ER+ BC risk. Therefore, replication across other populations or time periods presents a valuable avenue of future research.

No (or at most very weak) genetic correlations between life-history traits and ER+ BC risk taken together with previous phenotypic evidence of earlier ages at menarche, later ages at first birth, and fewer children increasing ER+ BC risk would also indirectly support the hypothesis that this is due to an evolutionary mismatch due to the rapid non-genetic changes in these traits over recent generations (Coe & Steadman, 1995; Eaton et al., 1994; Short, 1976; Strassmann, 1999). Changes in diet and nutrition have caused earlier ages at menarche (Duan et al., 2020; Nguyen et al., 2020; Rees, 1993) and recent changes in cultural preferences towards smaller families have driven later ages at first birth and fewer children (Aitken, 2022; Kirk, 1996). These changes may have led to higher lifetime exposure than a woman's cancer suppression mechanisms have evolved to handle (Coe & Steadman, 1995; Eaton et al., 1994; Short, 1976; Strassmann, 1999). In line with this hypothesis, differences in birth rates among countries predict breast cancer prevalence (You et al., 2018), with our findings of low heritability of these traits and low or no genetic

correlations further suggesting that these associations are environmental in origin. However, the extent to which changes in these life-history traits have also driven the increased prevalence in ER+ BC risk is unclear, and longitudinal studies would be best placed to answer.

Our study suffers from a few key limitations. First, we were not able to use family-based GWAS data for ER+ BC risk as to our knowledge no such studies exist, potentially biasing genetic correlation estimates. However, our estimates of genetic correlations between life-history traits and ER+ BC risk should remain relatively unbiased because: (1) we used family-based data for the life-history traits in all LD score regression analyses; and (2) our ER+ BC risk GWAS data suggested little evidence of bias due to factors such as population stratification (Supplementary note 1, Table S5). Further, estimating genetic correlations with GWAS data uncorrected for population stratification did not change our results qualitatively (Supplementary note 1, Table S6). Second, although we found little importance of genome-wide antagonistic pleiotropy for ER+ BC risk, exploring genetic correlations between estrogen receptor-negative breast cancer (ER- BC) risk and life-history traits may find different results. Interestingly, ER- BC risk shows higher heritability (Bauer et al., 2007), and there is some evidence that ER- BC risk is genetically correlated with earlier ages of reproduction (Prince et al., 2024), indicative of a fast life-history, as previously hypothesized (Hidaka & Boddy, 2016). Third, we were not able to account for age at menopause in our analyses due to insufficient sample size. Age at menopause is an important life-history trait and late ages at menopause are a risk factor for ER+ breast cancer (Kelsey & Bernstein, 1996) with previous studies showing some evidence for genetic variants underlying age at menopause also influencing cancer risk (Benonisdottir et al., 2024). Finally, despite using both individual-level biobank data and the largest available data of family-based GWAS data for life-history traits and GWAS data on ER+ BC risk, our genetic correlations (at least for age at menarche and first birth) were still somewhat imprecise. Thus, although our results suggest that non-genetic factors are the most important factor determining ER+ BC risk and life history traits, with greater power, future studies may be able to detect statistically significant genetic correlations between life-history traits and ER+ BC risk.

Our study provides an important step towards a more complete understanding of the genetic and evolutionary underpinnings of ER+ BC risk, which is a deadly form of cancer with increasing prevalence. Although our results suggest that the prevalence of ER+ BC risk variants in future generations is unlikely to change due to natural selection, we highlight the importance of examining the gene-by-environment interactions in the context of ER+ BC risk, which has so far received limited attention (but see Middha et al. 2023). This is an important avenue of research, with the broad range of non-genetic ER+ BC risk factors and the number of existing variants giving a high potential of gene-by-environment interactions. Finally,

we hope that our multifaceted approach can provide a template for future studies using the latest developments in genetic analyses to test evolutionary hypotheses.

METHODS

Lifelines biobank

Lifelines is a multi-disciplinary prospective population-based cohort study examining, in a unique three-generation design, the health and health-related behaviors of 167,729 persons living in the North of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics. We focused on data from 97,131 women born 1919-1979 and recruited 2007-2013 (https://data-catalogue.lifelines.nl/, accessed 08-05-2024, *Lifelines*, 2024).

Life-history traits

Age at menarche describes the age in which a female first started menstruating (Rees, 1993), can vary among individuals due to both environmental (e.g., nutritional status) and genetic effects, and is linked to future disease risk, including breast cancer (Parent et al., 2003; Perry et al., 2015). We quantified age at menarche using the lifelines baseline question: "How old were you when you had your first menstruation?" (N=85,210 women). For quality control, we removed 153 values >3 standard deviations from the mean. This gave a mean age at menarche of 13.1 (SD = 1.7, range = 8-18, N = 85,057, Figure S1A), representative of the Netherlands (Parent et al., 2003).

Age at first birth and number of children born are likewise influenced by both genetic factors and environmental factors which are then further constrained by the social environment, cultural background, and any remaining individual preferences (Barban et al., 2016; Mills et al., 2021). We quantified age at first birth and the number of children ever born using questions for a participant's birth date and the years of birth of her children focusing on full-term pregnancies. For these variables, we used data from baseline questionnaires, supplemented with follow-up questions during wave two of the Lifelines data collection (2014-2018). For age at first birth, we screened the questionnaire data removing any years of birth of children before an individual was born and after an individual had answered the questionnaire (n = 16). We used the birth dates of participants to calculate ages at first birth (mean = 26.9, 95% CIs = 18.9-36.3, Figure S1B) which occurred 1942-2017 (N = 64,779). The number of children ever born was then calculated only for women who were asked for the years of birth of their children after age 45 to avoid biasing the number

301 of childless women and the total number of children born. This gave a mean number of children born of 2.0
302 (range 0-8, N = 51,672, Figure S1C). All data were handled in R 4.2.2 (2022-10-31) using *dplyr* 1.1.4.

303 *ER+ BC diagnoses*

304 ER+ BC diagnoses were determined by linking Lifelines participants to the Dutch Nationwide Pathology
305 Databank (Palga) which contains the national records of histo- and cytopathology, with complete coverage
306 of the Netherlands since 1991 (www.palga.nl, (Casparie et al., 2007). The Palga databank details age and
307 dates of diagnoses, and pathology reports describing the subtype, nature, and grade of the tumor (Casparie
308 et al., 2007). Linkage of Lifelines participants to the Palga databank was successfully done for all, except
309 244 Lifelines participants, producing 10,114 unique pathology reports on breast tumors across 4,310
310 individuals.

311 We subtyped these records according to estrogen receptor status (ER-status) using the report conclusions
312 written by pathologists contained in the Palga excerpts (see Supplementary note 2 for further details).
313 According to national Dutch guidelines, ER-status is considered positive when >= 10% of tumor cells show
314 are ER+ (IKNL, 2012). We focused on women whose first cancer diagnosis was breast cancer to remove
315 possible metastatic tumors of different origin (n = 628). We also excluded 15 male diagnoses. This left 3,667
316 Lifelines participants with their first-cancer diagnosis being breast cancer, diagnosed from 1981-2023, and
317 2,177 ER+ diagnoses (Figure S2), 415 ER- diagnoses, with 1,090 being unable to be assigned an ER-status.
318 The ER+ BC diagnoses had a median age of diagnosis of 51 (range = 24-93),

319 *Genotype data*

320 Genotype data on European-Ancestry Lifelines participants were available on three different platforms: the
321 Illumina CytoSNP-12v2 array (hence CytoSNP array data), the Infinium Global Screening Array
322 MultiEthnic Disease Version 1.0 (hence Global screening array data), and the Affymetrix FinnGen Thermo
323 Fisher Axiom Custom array (hence Affymetrix array data). Full details about the quality control of each
324 genomic dataset and imputation can be found online at: https://wiki.lifelines.nl/. In brief, single nucleotide
325 polymorphisms (SNPs) were filtered according to minimum allele frequency, Hardy-Weinberg equilibrium
326 tests, Plink call rates, and ancestry outliers. Imputation was also performed using Minimac 2012.10.3
327 (Howie et al., 2012) with *The Genome of The Netherlands* release 5 (The Genome of the Netherlands
328 Consortium, 2014) and the 1000 Genomes phase1 v3 reference panels (The 1000 Genomes Project
329 Consortium, 2010).

330 We aimed to combine the different Lifelines genotypes in one GREML analysis and therefore harmonized
331 these three datasets using Genotype Harmonizer 1.4.25 (available at
332 https://github.com/molgenis/systemsgenetics/wiki/Genotype-Harmonizer). This software uses linkage

disequilibrium patterns to address strand issues, aligning genotype data to one reference dataset, for which we used the Global screening array. The harmonized data contained 1,032,001 variants across 80,045 individuals.

GREML analyses of Lifelines data

We performed GREML analyses (using GCTA/1.93.2beta-GCCcore-7.3.0, Yang et al., 2011) which leverages the genomic relatedness between distantly-related individuals in a population to examine how genetic variation associated with one trait covaries with another (Yang et al., 2010). We first performed a univariate GREML on ER+ BC risk, age at menarche, age at first birth, and number children of ever born to estimate the heritability (h^2) of each trait, which is the proportion of total phenotypic variance in a trait (σ_P^2) due to genetic factors (σ_G^2):

$$h^2 = \frac{\sigma_G^2}{\sigma_P^2}$$

With GREML, the heritability of a given phenotype (**y**) is estimated using a linear mixed model conditioned upon a vector of fixed effects (β) (such as age and sex) and where the variation explained by the genomic relatedness matrix (g) is divided by the residual variance (ε) to estimate the heritability in the equation:

$$y = X\beta + g + \varepsilon$$

with parameters estimated using the restricted maximum likelihood approach (see Yang et al., 2010, for comprehensive methods).

We then estimated the genetic correlation ($r(G)$) between ER+ BC risk and each life-history trait in three bivariate models (Lee et al., 2012). $r(G)$ is estimated as the genetic covariance between traits $Cov(G_{t1,t2})$ scaled by the genetic variance in trait:

$$r(G_{t1,t2}) = \frac{Cov(G_{t1,t2})}{\sqrt{V_{Gt1}} * \sqrt{V_{Gt2}}}$$

We created the genomic relatedness matrices (GRMs) for all individuals for which we had genomic data using the *--make-grm* argument from the *gcta64* function. We created GRMs for each chromosome first, then merged them into one GRM using *--mgrm*. Finally, the merged GRM was pruned of combinations of closely related individuals while maximizing the sample size using *--grm-cutoff* at a relatedness threshold of 0.05 (N=25,068) to remove bias in genetic variances due gene-environment correlations.

The age of the individual when they were either first diagnosed with breast cancer using the Palga databank, or (for individuals not diagnosed with ER+ BC) when they were last known to be alive and not have ER+

361 BC, was included a linear and squared covariate. The undiagnosed age could be the age of the individual at
362 the date the Palga records were up to date until (February 21st, 2023) or the date of death for an individual
363 if this occurred prior (n = 2,293). Dates of death were included in Lifelines as recorded by the Dutch
364 government and were known even in cases of emigration. These dates were decimalized using *lubridate*
365 1.9.3 in R and ages were calculated using the birth year and month available in Lifelines. We also included
366 the participant's birthdate as a linear and squared covariate, to account for any temporal trends in variables,
367 and the top 20 principal components of the pruned GRM used in each model to account for possible
368 population stratification (Yang et al., 2014). Principal components were generated using the *--pca* argument
369 from the *gcta64* function.

370 However, our bivariate models struggled to converge and univariate models provided apparently biased
371 heritability estimates inconsistent with the literature (e.g., 0% for age at menarche, Table S1d). We posited
372 that this was due to poor coverage of loci between the different genomic samples. For example, although
373 we had over 1 million SNPs in total, the number of SNPs measured in individuals belonging to, for example,
374 both the Affymetrix and CytoSNP array samples was ~2% of the total number of SNPs available (Table
375 S2). This could result in biased genomic relatedness estimates and increase the chance of missing causal
376 loci, downwardly biasing heritability estimates in the univariate models. We therefore created GRMs based
377 on each genomic dataset separately where the number of loci measured between individuals was larger and
378 more consistent (Table S2). We present the analysis using the Global screening array data in the main text
379 – because it was the largest in both number of SNPs and individuals – and use the analyses of the other
380 cohorts as replicates. Our sample sizes of informative individuals for our analyses using the Global
381 screening array were: age at menarche, 8,071; age at first birth, 6,260; number of children ever born, 4,602;
382 and 182 ER+ BC cases and 8,700 controls.

383 Likelihood ratio tests (LRTs) were used to test if heritabilities (in univariate models) and genetic correlations
384 (in bivariate models) were statistically different from 0. For genetic correlations, we adjusted for multiple
385 testing across our three life-history traits using Bonferroni correction to a cutoff of 0.05, giving an adjusted
386 significance threshold of $0.01\dot{6}$. To aid convergence, we also used the Expectation–maximization algorithm,
387 increased the number of iterations to 10,000, and standardized life-history trait values to a mean of zero and
388 standard deviation of one.

389 *LD score regression of GWAS data*

390 We then replicated our findings from the genomic relatedness analyses using LD score regression to estimate
391 heritabilities and genetic correlations between our traits from previously published GWAS summary
392 statistics (Bulik-Sullivan, Loh, et al., 2015). LD score regression combines genome-wide association study

(GWAS) data (i.e., statistical associations between all genetic variants and a trait from a particular study) across different analyses to examine how strongly correlated the effects of genetic variants are across different traits, controlling for their genetic linkage. For ER+ BC risk, we used GWAS summary statistics from Michailidou et al., (2017, available at: https://www.ebi.ac.uk/gwas/studies/GCST004988) which contained associations of 11,792,542 SNPs with ER+ BC risk based upon 69,501 cases and 95,042 controls (see Michailidou et al., 2017, for full methods). We cleaned these data using the *munge_sumstats* function from the software using *ldsc* 1.0.1 (Bulik-Sullivan, Loh, et al., 2015) supported by *Miniconda3*. Specifically, we removed SNPs not found in the HapMap3 to control for imputation quality in the absence of imputation quality scores (n = 10,643,464), with MAF < 0.01 (n = 10,735), missing values (n = 8,108), that were strand ambiguous (n = 379), and duplicated identifiers (n = 408). This left 1,129,448 SNPs with a mean χ^2 value of 1.537, 2,217 of which held genome-wide significance.

For our life-history traits, we used SNP summary statistics from Tan et al. (2024): a recent family-based GWAS study which aimed to separate direct genetic effects of a given SNP on a trait from indirect genetic effects and factors such as assortative mating and population stratification (Kong et al., 2018; Young et al., 2022). This can be achieved by conditioning genetic associations found across individuals upon the parental genotypes in the GWAS model regression; see Tan et al., (2024) for a full description of methods.

We downloaded the Tan et al., (2024) summary statistics from the SSGAC data portal (https://thessgac.com/), which contained the summary statistics of 5,671,230 SNPs for age at menarche, 6,551,047 for age at first birth, and 6,412,325 for number of children ever born. We again used the *munge_sumstats* function to clean the data prior to analyses and, again, selected only cites used in HapMap3, and – as advised by Tan et al., (2024) – filtered for SNPs with an effective sample size greater than the median for each trait to ensure summary statistics were included for only loci whose genetic effects were well estimated. The number of SNPs with known associations for each trait after cleaning were 580,219 for age at menarche, 708,027 for age at first birth, and 670,251 for number of children ever born. The mean and number of genome-wide significant SNPs for these traits were 1.127 and 26, 1.078 and 0, and 1.051 and 0, respectively.

Heritabilities and genetic correlations were estimated using the software *ldsc* (1.0.1, Bulik-Sullivan, Finucane, et al., 2015; Bulik-Sullivan, Loh, et al., 2015). *ldsc* estimates the heritability of a trait by regressing the χ^2-statistics from SNPs on linkage disequilibrium scores using the equation:

$$E\left[\chi_j^2\right] \approx \frac{N_j h_g^2}{M} l_j + 1$$

where $E[\chi_j^2]$ is the expected χ^2-statistics for the association between the trait and SNP j, N_j is the number of individuals measured for the trait for SNP j, h_g^2 is the squared correlation between the phenotype and the best linear unbiased predictor estimated from the SNPs, M is the total number of SNPs, and l_j denotes the LD score of SNP j (Bulik-Sullivan, Loh, et al., 2015). In all models, we used European ancestry LD scores from the 1000 Genome project (The 1000 Genomes Project Consortium, 2010) available from online at: https://zenodo.org/records/7768714. We removed loci in the major histocompatibility complex because of its unusual genetic architecture.

Genome-wide genetic correlations were estimated using the following equation where the expected value of $z_{1j}z_2l_j$ for a SNP j is

$$E[z_{1j}z_2l_j] = \frac{\sqrt{N_1N_2}\varrho_g}{M}l_j + \frac{\varrho N_s}{\sqrt{N_1N_2}}$$

where N_i is the sample size for the study i, ϱ_g is the genetic covariance, l_j is the LD score, N_s is the number of total individuals in both studies, ϱ is the phenotypic correlation among the overlapping samples (Bulik-Sullivan, Finucane, et al., 2015). We can then normalize the genetic covariance using the heritabilities from trait 1, h_1^2, and trait 2, h_2^2, to estimate the genetic correlation, r_g, with the equation:

$$r_g = \frac{\varrho_g}{\sqrt{h_1^2h_2^2}}$$

We used the same LD scores and GWAS data as above. The sample size for SNPs with available LD scores, and GWAS summary statistics for both traits for our three bivariate models were: age at menarche = 551,453, age at first birth = 668,389, and number of children ever born = 635,154. Significance of the genetic correlations was determined by z-score derived p-values provided by *ldsc,* adjusted to a significance threshold of $0.01\dot{6}$. We also report the heritabilities estimated for each trait in the genetic correlation models and the mean z-scores of each SNPs multiplied association with both traits. As a supplementary analysis we also estimated our genetic correlations using the genetic effects from Tan *et al.* (2024) that did not correct for environmental confounding (i.e., population genetic effects), but this did not produce qualitatively different results (Table S6).

DATA AVAILABILITY

Data may be obtained from Lifelines and are not publicly available. More information about how to request Lifelines data and the conditions of use can be found on their website (https://www.lifelines-biobank.com/researchers/working-with-us/step-1-prepare-and-submit-your-application).

FUNDING

The Lifelines initiative has been made possible by subsidy from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen (UMCG), Groningen University and the Provinces in the North of the Netherlands (Drenthe, Friesland, Groningen). E.A.Y.'s PhD was funded by the University of Groningen, through a Rosalind Franklin Fellowship awarded to H.L.D.

ACKNOWLEDGEMENTS

We thank the Lifelines data management team and especially Sylvia Gerritsma for their support during this project and Mariana Escobar Rodríguez, Grigory Sivolenko, Thijs Janzen, and Raphaël Scherrer for their advice on the data handling. Finally, we would also like to thank the University of Exeter's quantitative genetics club and Luca G. Hahn for conceptual feedback and discussions regarding the results.

ETHICAL APPROVAL INFORMATION

The Lifelines Cohort Study is conducted following the principles of the Declaration of Helsinki and with approval from the Medical Ethics Committee of the University Medical Center Groningen, Groningen (number 2007/152). Written informed consent was given by each Lifelines participant and the research presented here were approved by Lifelines board (number OV17_00378), adhering to the Lifelines User Code of Conduct. All data was pseudo-anonymized before being received and was handled in accordance with the General Data Protection and Regulation act. This research was approved by the scientific commission of Palga (number 2023-126). Further, data were accessed only within the secure Lifelines workspace or cluster environment on the University Medical Center Groningen cluster *Gearshift* (https://docs.gcc.rug.nl/gearshift/cluster/).

REFERENCES

Aitken, R. J. (2022). The changing tide of human fertility. *Human Reproduction*, *37*(4), 629–638. https://doi.org/10.1093/humrep/deac011

Aktipis, C. A., Boddy, A. M., Jansen, G., Hibner, U., Hochberg, M. E., Maley, C. C., & Wilkinson, G. S. (2015). Cancer across the tree of life: Cooperation and cheating in multicellularity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1673), 20140219. https://doi.org/10.1098/rstb.2014.0219

Aktipis, C. A., Ellis, B. J., Nishimura, K. K., & Hiatt, R. A. (2015). Modern reproductive patterns associated with estrogen receptor positive but not negative breast cancer susceptibility. *Evolution, Medicine and Public Health*, *2015*(1), 52–74. https://doi.org/10.1093/emph/eou028

Barban, N., Jansen, R., De Vlaming, R., Vaez, A., Mandemakers, J. J., Tropf, F. C., Shen, X., Wilson, J. F., Chasman, D. I., Nolte, I. M., Tragante, V., van Der Laan, S. W., Perry, J. R. B., Kong, A., Ahluwalia, T. S., Albrecht, E., Yerges-Armstrong, L., Atzmon, G., Auro, K., … Mills, M. C. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature Genetics*, *48*(12), 1462–1472. https://doi.org/10.1038/ng.3698

Benonisdottir, S., Straub, V. J., Kong, A., & Mills, M. C. (2024). Genetics of female and male reproductive traits and their relationship with health, longevity and consequences for offspring. *Nature Aging*, *4*(12), 1745–1759. https://doi.org/10.1038/s43587-024-00733-w.

Berg, J. J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A. M., Mostafavi, H., Field, Y., Boyle, E. A., Zhang, X., Racimo, F., Pritchard, J. K., & Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *eLife*, *8*, e39725. https://doi.org/10.7554/eLife.39725

Bieuville, M., Faugère, D., Galibert, V., Henard, M., Dujon, A. M., Ujvari, B., Pujol, P., Roche, B., & Thomas, F. (2023). Number of lifetime menses increases breast cancer occurrence in postmenopausal women at high familial risk. *Frontiers in Ecology and Evolution*, *11*, 912083. https://doi.org/10.3389/fevo.2023.912083.

Bonnet, T., Morrissey, M. B., Villemereuil, P. D., Alberts, S. C., Arcese, P., Bailey, L. D., Boutin, S., Brekke, P., Brent, L. J. N., Camenisch, G., Charmantier, A., Clutton-Brock, T. H., Cockburn, A., Coltman, D. W., Courtiol, A., Davidian, E., Evans, S. R., Ewen, J. G., Festa-bianchet, M., … Kruuk, L. E. B. (2022). Genetic variance in fitness indicates rapid on-going adaptive evolution in wild animals. *Science*, *376*(6596), 1012–1016.

Border, R., Athanasiadis, G., Buil, A., Schork, A. J., Cai, N., Young, A. I., Werge, T., Flint, J., Kendler, K. S., Sankararaman, S., Dahl, A. W., & Zaitlen, N. A. (2022). Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science*, *378*(6621), 754761.

Bulik-Sullivan, B. K., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the

512 Wellcome Trust Case Control Consortium 3, Duncan, L., Perry, J. R. B., Patterson, N., Robinson,
513 E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across
514 human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406
515 Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Consortium, S. W. G. of the P. G.,
516 Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes
517 confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–
518 295. https://doi.org/10.1038/ng.3211
519 Byars, S. G., Ewbank, D., Govindaraju, D. R., & Stearns, S. C. (2010). Natural selection in a
520 contemporary human population. *Proceedings of the National Academy of Sciences of the United*
521 *States of America*, *107*(SUPPL. 1), 1787–1792. https://doi.org/10.1073/pnas.0906199106
522 Byars, S. G., & Voskarides, K. (2020). Antagonistic Pleiotropy in Human Disease. *Journal of Molecular*
523 *Evolution*, *88*(1), 12–25. https://doi.org/10.1007/s00239-019-09923-2
524 Casparie, M., Tiebosch, A. T. M. G., Burger, G., Blauwgeers, H., van De Pol, A., van Krieken, J. H. J. M.,
525 & Meijer, G. A. (2007). Pathology Databanking and Biobanking in The Netherlands, a Central
526 Role for PALGA, the Nationwide Histopathology and Cytopathology Data Network and Archive.
527 *Analytical Cellular Pathology*, *29*(1), 19–24. https://doi.org/10.1155/2007/971816
528 Charlesworth, B. (1980). *Evolution in age-structured populations*. Cambridge University Press
529 Cambridge.
530 Charlesworth, B. (2001). Patterns of age-specific means and genetic variances of mortality rates predicted
531 by the mutation-accumulation theory of ageing. *Journal of Theoretical Biology*, *210*(1), 47–65.
532 Coe, K., & Steadman, L. B. (1995). The human breast and the ancestral reproductive cycle: A preliminary
533 inquiry into breast cancer etiology. *Human Nature*, *6*(3), 197–220.
534 https://doi.org/10.1007/BF02734139
535 Colditz, G. A., Rosner, B. A., Chen, W. Y., Holmes, M. D., & Hankinson, S. E. (2004). Risk Factors for
536 Breast Cancer According to Estrogen and Progesterone Receptor Status. *JNCI Journal of the*
537 *National Cancer Institute*, *96*(3), 218–228. https://doi.org/10.1093/jnci/djh025
538 Crespi, B. J., & Summers, K. (2006). Positive selection in the evolution of cancer. *Biological Reviews of*
539 *the Cambridge Philosophical Society*, *81*(3), 407–424.
540 https://doi.org/10.1017/S1464793106007056
541 Dall, G. V., Hawthorne, S., Seyed-Razavi, Y., Vieusseux, J., Wu, W., Gustafsson, J.-A., Byrne, D.,
542 Murphy, L., Risbridger, G. P., & Britt, K. L. (2018). Estrogen receptor subtypes dictate the
543 proliferative nature of the mammary gland. *Journal of Endocrinology*, *237*(3), 323–336.
544 https://doi.org/10.1530/JOE-17-0582

Day, F. R., Thompson, D. J., Helgason, H., Chasman, D. I., Finucane, H., Sulem, P., Ruth, K. S., Whalen, S., Sarkar, A. K., Albrecht, E., Altmaier, E., Amini, M., Barbieri, C. M., Boutin, T., Campbell, A., Demerath, E., Giri, A., He, C., Hottenga, J. J., … Smith, G. D. (2017). Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nature Genetics*, *49*(6), 834–841. https://doi.org/10.1038/ng.3841

Duan, R., Chen, Y., Qiao, T., Duan, R., Chen, M., Zhao, L., Gong, Y., & Cheng, G. (2020). Modern dietary pattern is prospectively associated with earlier age at menarche: Data from the CHNS 1997–2015. *Nutrition Journal*, *19*(1), 95. https://doi.org/10.1186/s12937-020-00622-z

Dujon, A. M., Boutry, J., Tissot, S., Lemaître, J.-F., Boddy, A. M., Gérard, A.-L., Alvergne, A., Arnal, A., Vincze, O., Nicolas, D., Giraudeau, M., Telonis-Scott, M., Schultz, A., Pujol, P., Biro, P. A., Beckmann, C., Hamede, R., Roche, B., Ujvari, B., & Thomas, F. (2022). Cancer Susceptibility as a Cost of Reproduction and Contributor to Life History Evolution. *Frontiers in Ecology and Evolution*, *10*(May), 1–18. https://doi.org/10.3389/fevo.2022.861103

Eaton, S. B., Pike, M. C., Short, R. V., Lee, N. C., Trussell, J., Hatcher, R. A., Wood, J. W., Worthman, C. M., Konner, M. J., Hill, K. R., Bailey, R., & Hurtado, A. M. (1994). Women's Reproductive Cancers in Evolutionary Context. *The Quarterly Review of Biology*, *69*(3), 353–367. https://doi.org/10.1086/418650

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford University Press.

GBD. (2021). *Global Burden of Disease Study 2021* [Dataset]. Global Burden of Disease Collaborative Network, Institute for Health Metrics and Evaluation (IHME). https://vizhub.healthdata.org/gbd-results/

Haldane, J. (1941). *New paths in genetics.* Allen & Unwin.

Haldane, J. (1954). *The measurement of natural selection*. *1*, 480–487.

Hidaka, B. H., & Boddy, A. M. (2016). Is estrogen receptor negative breast cancer risk associated with a fast life history strategy? *Evolution, Medicine and Public Health*, *2016*(1), 17–20. https://doi.org/10.1093/EMPH/EOV034

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., Chittoor, G., Ahlskog, R., Lind, P. A., Palviainen, T., van Der Zee, M. D., Cheesman, R., Mangino, M., Wang, Y., Li, S., Klaric, L., Ratliff, S. M., Bielak, L. F., Nygaard, M., … Davies, N. M. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics*, *54*(5), 581–592. https://doi.org/10.1038/s41588-022-01062-7

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*, *44*(8), 955–959. https://doi.org/10.1038/ng.2354

IKNL. (2012). *Breast Cancer Guideline*. The Netherlands Comprehensive Cancer Organization. https://www.nabon.nl/wp-content/uploads/2022/10/Dutch-Breast-Cancer-Guideline-2012.pdf

Jacqueline, C., Biro, P. A., Beckmann, C., Moller, A. P., Renaud, F., Sorci, G., Tasiemski, A., Ujvari, B., & Thomas, F. (2017). Cancer: A disease at the crossroads of trade-offs. *Evolutionary Applications*, *10*(3), 215–225. https://doi.org/10.1111/eva.12444

Jiang, X., Finucane, H. K., Schumacher, F. R., Schmit, S. L., Tyrer, J. P., Han, Y., Michailidou, K., Lesseur, C., Kuchenbaecker, K. B., Dennis, J., Conti, D. V., Casey, G., Gaudet, M. M., Huyghe, J. R., Albanes, D., Aldrich, M. C., Andrew, A. S., Andrulis, I. L., Anton-Culver, H., … Lindström, S. (2019). Shared heritability and functional enrichment across six solid cancers. *Nature Communications*, *10*(1), 431. https://doi.org/10.1038/s41467-018-08054-4

Kapsetaki, S. E., Compton, Z. T., Dolan, J., Harris, V. K., Mellon, W., Rupp, S. M., Duke, E. G., Harrison, T. M., Aksoy, S., Giraudeau, M., Vincze, O., McGraw, K. J., Aktipis, A., Tollis, M., Boddy, A. M., & Maley, C. C. (2024). Life history traits and cancer prevalence in birds. *Evolution, Medicine, and Public Health*, *12*(1), 105–116. https://doi.org/10.1093/emph/eoae011

Kelsey, J. L., & Bernstein, L. (1996). Epidemiology and prevention of breast cancer. *Annual Review of Public Health*, *17*(1), 47–67.

Kirk, D. (1996). Demographic transition theory. *Population Studies*, *50*(3), 361–387.

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdottir, S., Oddsson, A., Halldorsson, B. V., Masson, G., & others. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424–428.

Lande, R., & Arnold, S. J. (1983). The Measurement of Selection on Correlated Characters. *Evolution*, *37*(6), 1201226. https://www.jstor.org/stable/2408842

Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, *28*(19), 2540–2542. https://doi.org/10.1093/bioinformatics/bts474

Lifelines. (2024). [Dataset]. https://data-catalogue.lifelines.nl/

Medawar, P. (1952). An unsolved problem in biology. Delineates of Theory of Mutation Accumulation. *Lewis, HK London, UK. Reprinted in: Medawar, PB (1957). The Uniqueness of the Individual. Methuen. London, UK*, 44–70.

Merlo, L. M. F., Pepper, J. W., Reid, B. J., & Maley, C. C. (2006). Cancer as an evolutionary and ecological process. *Nature Reviews Cancer*, *6*(12), 924–935. https://doi.org/10.1038/nrc2013

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., Bolla, M. K., Wang, Q., Tyrer, J., Dicks, E., Lee, A., Wang, Z., Allen, J.,

613 Keeman, R., Eilber, U., … Easton, D. F. (2017). Association analysis identifies 65 new breast
614 cancer risk loci. *Nature*, *551*(7678), 92–94. https://doi.org/10.1038/nature24284
615 Michailidou, K., The Breast and Ovarian Cancer Susceptibility Collaboration, Hereditary Breast and
616 Ovarian Cancer Research Group Netherlands (HEBON), kConFab Investigators, Australian
617 Ovarian Cancer Study Group, The GENICA (Gene Environment Interaction and Breast Cancer in
618 Germany) Network, Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L.,
619 Schmidt, M. K., Chang-Claude, J., Bojesen, S. E., Bolla, M. K., Wang, Q., Dicks, E., Lee, A.,
620 Turnbull, C., … Easton, D. F. (2013). Large-scale genotyping identifies 41 new loci associated
621 with breast cancer risk. *Nature Genetics*, *45*(4), 353–361. https://doi.org/10.1038/ng.2563
622 Middha, P., Wang, X., Behrens, S., Bolla, M. K., Wang, Q., Dennis, J., Michailidou, K., Ahearn, T. U.,
623 Andrulis, I. L., Anton-Culver, H., Arndt, V., Aronson, K. J., Auer, P. L., Augustinsson, A., Baert,
624 T., Freeman, L. E. B., Becher, H., Beckmann, M. W., Benitez, J., … Chang-Claude, J. (2023). A
625 genome-wide gene-environment interaction study of breast cancer risk for women of European
626 ancestry. *Breast Cancer Research*, *25*(1), 93. https://doi.org/10.1186/s13058-023-01691-8
627 Mills, M. C., Tropf, F. C., Brazel, D. M., van Zuydam, N., Vaez, A., Agbessi, M., Ahsan, H., Alves, I.,
628 Andiappan, A. K., Arindrarto, W., Awadalla, P., Battle, A., Beutner, F., Jan Bonder, M.,
629 Boomsma, D. I., Christiansen, M. W., Claringbould, A., Deelen, P., Esko, T., … Day, F. R.
630 (2021). Identification of 371 genetic variants for age at first sex and birth linked to externalising
631 behaviour. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01135-3
632 Momenimovahed, Z., & Salehiniya, H. (2019). Epidemiological characteristics of and risk factors for
633 breast cancer in the world. *Breast Cancer: Targets and Therapy*, *Volume 11*, 151–164.
634 https://doi.org/10.2147/BCTT.S176070
635 Nguyen, N., Fan, H.-Y., Tsai, M.-C., Tung, T.-H., Huynh, Q., Huang, S.-Y., & Chen, Y. (2020). Nutrient
636 Intake through Childhood and Early Menarche Onset in Girls: Systematic Review and Meta-
637 Analysis. *Nutrients*, *12*(9), 2544. https://doi.org/10.3390/nu12092544
638 Nunney, L. (2013). The real war on cancer: The evolutionary dynamics of cancer suppression.
639 *Evolutionary Applications*, *6*(1), 11–19. https://doi.org/10.1111/eva.12018
640 Parent, A.-S., Teilmann, G., Juul, A., Skakkebaek, N. E., Toppari, J., & Bourguignon, J.-P. (2003). The
641 Timing of Normal Puberty and the Age Limits of Sexual Precocity: Variations around the World,
642 Secular Trends, and Changes after Migration. *Endocrine Reviews*, *24*(5), 668–693.
643 https://doi.org/10.1210/er.2002-0019
644 Perry, J. R. B., Murray, A., Day, F. R., & Ong, K. K. (2015). Molecular insights into the aetiology of
645 female reproductive ageing. *Nature Reviews Endocrinology*, *11*(12), 725–734.
646 https://doi.org/10.1038/nrendo.2015.167

Prince, C., Howe, L. D., Sanderson, E., Sharp, G. C., Fraser, A., Lloyd-Lewis, B., & Richmond, R. C. (2024). Evaluating the causal impact of reproductive factors on breast cancer risk: A multivariable mendelian randomization approach. *medRxiv*. https://doi.org/10.1101/2024.02.02.24301815

Promislow, D. E. L., & Harvey, P. H. (1990). Living fast and dying young: A comparative analysis of life-history variation among mammals. *Journal of Zoology*, *220*(3), 417–437. https://doi.org/10.1111/j.1469-7998.1990.tb04316.x

Rees, M. (1993). Menarche when and why? *The Lancet*, *342*(8884), 1375–1376. https://doi.org/10.1016/0140-6736(93)92746-G

Robertson, A. (1966). A mathematical model of the culling process in dairy cattle. *Animal Science*, *8*(1), 95–108.

Rodríguez, J. A., Marigorta, U. M., Hughes, D. A., Spataro, N., Bosch, E., & Navarro, A. (2017). Antagonistic pleiotropy and mutation accumulation influence human senescence and disease. *Nature Ecology and Evolution*, *1*(3), 1–5. https://doi.org/10.1038/s41559-016-0055

Roff, D. A. (1992). *The evolution of life histories: Theory and analysis*. Chapman & Hall; WorldCat.

Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J., Dotinga, A., Vonk, J. M., van Dijk, F., van Zon, S. K., Wijmenga, C., Wolffenbuttel, B. H., & Stolk, R. P. (2015). Cohort Profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, *44*(4), 1172–1180. https://doi.org/10.1093/ije/dyu229

Short, R. V. (1976). The evolution of human reproduction. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *195*(1118, A Discussion on Contraceptives of the Future), 3–24. https://doi.org/10.1098/rspb.1976.0095

Smith, K. R., Hanson, H. A., Mineau, G. P., & Buys, S. S. (2012). Effects of BRCA1 and BRCA2 mutations on female fertility. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1732), 1389–1395. https://doi.org/10.1098/rspb.2011.1697

Stearns, S. C. (1977). The Evolution of Life History Traits: A Critique of the Theory and a Review of the Data. *Annual Review of Ecology and Systematics*, *8*(1), 145–171. https://doi.org/10.1146/annurev.es.08.110177.001045

Stearns, S. C. (1992). *The evolution of life histories*. Oxford University Press, Oxford.

Strassmann, B. I. (1999). Menstrual Cycling and Breast Cancer: An Evolutionary Perspective. *Journal of Women's Health*, *8*(2), 193–202. https://doi.org/10.1089/jwh.1999.8.193

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Tan, T., Jayashankar, H., Guan, J., Nehzati, S. M., Mir, M., Bennett, M., Agerbo, E., Ahlskog, R., Pinto De Andrade Anapaz, V., Asvold, B. O., Benonisdottir, S., Bhatta, L., Boomsma, D. I., Brumpton, B., Campbell, A., Chabris, C. F., Cheesman, R., Chen, Z., China Kadoorie Biobank Collaborative Group, … Young, A. S. (2024). *Family-GWAS reveals effects of environment and mating on genetic associations*. https://doi.org/10.1101/2024.10.01.24314703

The 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061–1073. https://doi.org/10.1038/nature09534

The Genome of the Netherlands Consortium. (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature Genetics*, *46*(8), 818–825. https://doi.org/10.1038/ng.3021

Trumble, B. C., Charifson, M., Kraft, T., Garcia, A. R., Cummings, D. K., Hooper, P., Lea, A. J., Rodriguez, D. E., Koebele, S. V., Buetow, K., Beheim, B., Minocher, R., Gutierrez, M., Thomas, G. S., Gatz, M., Stieglitz, J., Finch, C. E., Kaplan, H., & Gurven, M. (2023). Apolipoprotein-ε4 is associated with higher fecundity in a natural fertility population. *SCIENCE ADVANCES*.

Williams, G. C. (1957). Pleiotropy, Natural Selection, and the Evolution of Senescence. *Evolution*, *11*(4), 398–411. https://www.jstor.org/stable/2406060

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., & Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569. https://doi.org/10.1038/ng.608

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *The American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, *46*(2), 100–106. https://doi.org/10.1038/ng.2876

Yasui, Y., & Potter, J. D. (1999). The shape of age–incidence curves of female breast cancer by hormone-receptor status. *Cancer Causes & Control*, *10*, 431–437.

You, W., Symonds, I., Rühli, F. J., & Henneberg, M. (2018). Decreasing Birth Rate Determining Worldwide Incidence and Regional Variation of Female Breast Cancer. *Advances in Breast Cancer Research*, *07*(01), 1–14. https://doi.org/10.4236/abcr.2018.71001

Young, A. I., Nehzati, S. M., Benonisdottir, S., Okbay, A., Jayashankar, H., Lee, C., Cesarini, D., Benjamin, D. J., Turley, P., & Kong, A. (2022). Mendelian imputation of parental genotypes

714 improves estimates of direct genetic effects. *Nature Genetics*, *54*(6), 897–905.
715 https://doi.org/10.1038/s41588-022-01085-0
716 Zhang, H., Ahearn, T. U., Lecarpentier, J., Barnes, D., Beesley, J., Qi, G., Jiang, X., O'Mara, T. A., Zhao,
717 N., Bolla, M. K., Dunning, A. M., Dennis, J., Wang, Q., Ful, Z. A., Aittomäki, K., Andrulis, I. L.,
718 Anton-Culver, H., Arndt, V., Aronson, K. J., … García-Closas, M. (2020). Genome-wide
719 association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-
720 specific analyses. *Nature Genetics*, *52*(6), 572–581. https://doi.org/10.1038/ng.1049
721