Genetic variation in reproductive 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], 0000-0002-2128-7587

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. Both evolutionary life-history theory and evidence studying individual ER+ BC risk variants suggest that 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, but how well this pattern is replicated across the full polygenic architecture of these complex traits is unclear. Here, we estimate genome-wide genetic correlations between ER+ BC risk and three reproductive 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 performing linkage disequilibrium score regression on European population and family-based genome-wide association study data. Regardless of the data or method used, genetic correlations were low and not statistically different from zero. These results therefore suggest that while some important loci underlying ER+ BC risk may demonstrate antagonistic pleiotropy symptomatic of life-history trade-offs, the available genomic data suggests these findings are not replicated, on average, across the genome. Consequently, our results suggest that genetic variation does not overall shape the observed phenotypic relationships between these life-history traits and ER+ BC risk. Future studies incorporating local-genetic correlations or gene-network data could further our understanding how as to how genetic variation in these traits maps to phenotypic associations and aid our understanding of the evolutionary processes shaping this important disease.

KEYWORDS: life-history trade-offs, genetic correlations, evolutionary medicine, quantitative genetics, SNP heritability, antagonistic pleiotropy, GWAS data

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 variation promoting a life-threatening disease persists in contemporary human populations. Examining how genetic variation in ER+ BC risk is associated with human life-history traits would give insight into the evolutionary processes maintaining ER+ BC risk variants, the disease's overall genetic architecture, and help quantify the extent to which these risk variants will become more or less prevalent in future generations.

Evolutionary life-history theory and the antagonistic pleiotropy theory of aging hypothesize that genetic risk variants for diseases such as cancer may persist in populations due to their beneficial effects at younger ages, as a part of life-history trade-offs (Jacqueline et al., 2017; Promislow & Harvey, 1990; Roff, 1992; Stearns, 1977, 1992; Williams, 1957). For example, genetic variants 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 (Figure 1, Dujon et al., 2022; Jacqueline et al., 2017). Indeed, although 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), cancer remains common across the tree of life suggesting the presence of evolutionary constraints (Aktipis et al., 2013; Boddy et al., 2015), with a recent study also suggesting the presence of a reproduction-cancer suppression trade-off amongst bird species (Kapsetaki et al., 2024). However, whether such life-history trade-offs also shape genetic risk for ER+ BC is unclear.

In humans, the genetics of reproductive traits can play an important role in health and disease (reviewed in Benonisdottir et al., 2024), including genetic predispositions to earlier reproduction and higher numbers of children being associated with faster aging (Long & Zhang, 2023; Wang et al., 2013) and increased risk for specific diseases [(Byars et al., 2017; Trumble et al., 2023) reviewed in (Byars & Voskarides, 2020)]. Cancer appears to be no exception, with a host of tumor suppressor genes also being linked to the regulation of the timing of puberty and fertility (Boddy et al., 2015; Byars & Voskarides, 2020; Crespi & Summers, 2006; Dujon et al., 2022; Jacqueline et al., 2017). These include: *p53*, where genetic variants increasing cancer risk also appear to increase conception likelihoods (Kang et al., 2009); and more repeats

of the *CAG* repeat region of the androgen receptor being associated with both increased fertility and higher prostate cancer risk (Summers & Crespi, 2008). Breast cancer (and in particular ER+ BC) is a good candidate for antagonistically pleiotropic genes maintained by life-history trade-offs considering that ER+ breast cancer and these life-history traits share hormonal pathways underpinned by genes (Ghosh et al., 2020; Jacqueline et al., 2017; Jasienska et al., 2017). One genealogical study from Utah found that carriers of *BRCA1/2* mutations also had more children (Smith et al., 2012) suggesting that these trade-offs may have shaped the evolution of some of the most important breast cancer risk variants. More contemporary studies examining all genome-wide significant loci for age at menarche, found that genetic variants lowering age at menarche were also associated with higher overall breast cancer risk (Xiang et al., 2025) and, specifically, higher ER+ BC risk (Day et al., 2017; Zhao et al., 2024).

Yet, these associations between human life-history traits and ER+ BC risk expected from life-history theory (and evidenced in genetic studies) are not entirely consistent with the phenotypic associations from epidemiological studies. While earlier ages at menarche have been associated phenotypically with higher ER+ BC risk, earlier ages at first birth and higher numbers of children are actually associated with lower ER+ BC risk (Figure 1, Aktipis et al., 2015; Kelsey et al., 1993; Momenimovahed & Salehiniya, 2019). This may be due to these trends resulting in women spending less time in amenorrhea – which is induced by breastfeeding – and therefore experiencing more menstrual cycles and having a higher lifetime exposure to estrogen (Eaton et al., 1994; Short, 1976; Strassmann, 1999). Estrogen in turn 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 also suggests that ER+ BC risk variants are – contrary to life-history theory and genetic evidence – under negative selection in contemporary populations. These contrasting findings are not mutually exclusive due to the importance 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) which may mask underlying genetic trade-offs. Moreover, previous genetic studies have largely focused on genome-wide significant variants which fail to capture the full scope of genetic variation in both ER+ BC risk and human life-history traits which are polygenic traits. Specifically, genome-wide significant variants explain less than 50% of the total variance in genomic risk for ER+ BC (Zhang et al., 2020), amount to only 25% of the total genetic variation in age at menarche (Day et al., 2017), with even fewer genome-wide significant variants for age at first birth and number of children born being detected despite significant heritability for both traits (Barban et al., 2016; Mills et al., 2021; Tropf et al., 2015).

To help understand these inconsistent results, we need to estimate genetic correlations using whole-genome data. Genetic correlations estimated from genomic data reflect the proportion of genetic variance

shared by traits and are a measure of the overall contribution of pleiotropy to the covariation of these traits (van Rheenen et al., 2019). We can therefore estimate genetic correlations between ER+ BC risk and human life-history traits to quantify to what extent life-history trade-offs shape ER+ BC risk on average across all loci, whether these trade-offs contribute genetically to the observed phenotypic trends, or whether observed phenotypic associations reflect underlying genetic factors (Figure 1). To the best of our knowledge, 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. This suggests that these phenotypic trends do not translate into evolutionary change (Haldane, 1954; Lande & Arnold, 1983; Robertson, 1966). However, these studies did not fully 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 shape the prevalence of these genetic variants.

To estimate genetic correlations between human reproductive life-history traits and ER+ BC risk we first perform a genomic restricted maximum likelihood (GREML) analysis (Lee et al., 2012; Yang et al., 2010) using the Dutch biobank Lifelines (Scholtens et al., 2015) linked to national pathology records (Casparie et al., 2007). We then repeat our analyses by estimating genetic correlations using linkage disequilibrium (LD) score regressions of the largest available GWAS data (Bulik-Sullivan, Finucane, et al., 2015). We first use 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). We then repeat our LD score regression analysis using population GWAS data as family-based GWAS data can suffer from reduced power and can underestimate genetic effects in the presence of assortative mating (Kemper et al., 2021).

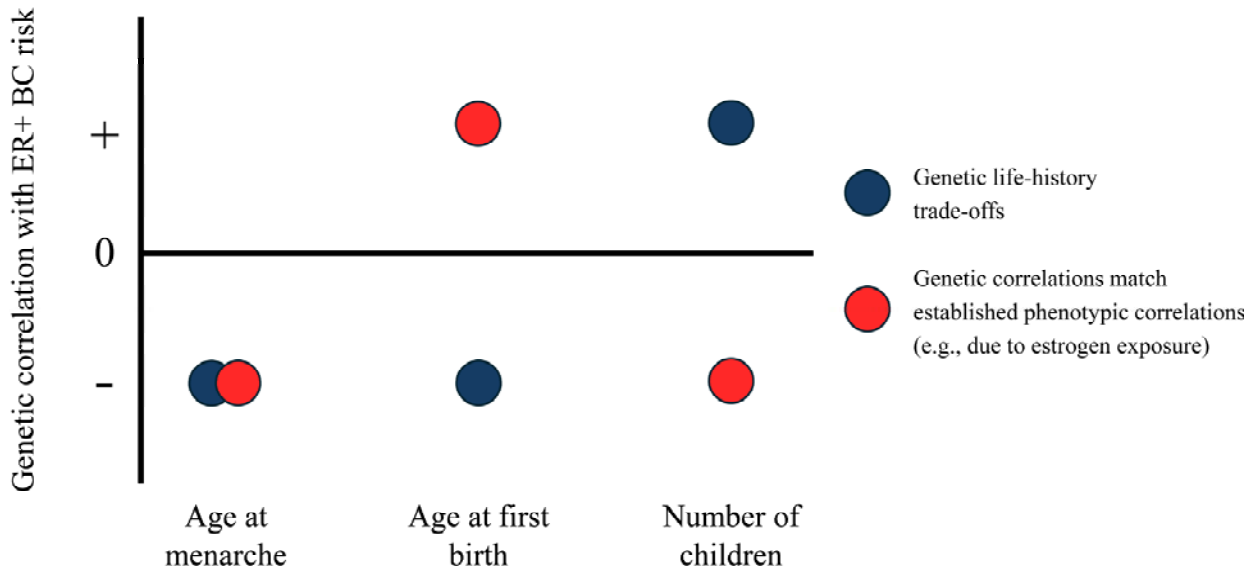


Figure 1: Predicted direction of genetic correlations between ER+ BC risk and reproductive life-history traits (age at menarche, age at first birth, and number of children) under different hypotheses. Circles below the x-axis indicate negative genetic correlations, i.e. genetic predispositions for lower/earlier 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: *genetic life-history trade-offs* as predicted by antagonistic pleiotropy theory of ageing and life-history 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 of Lifelines data

We first performed GREML analyses on data from the Lifelines biobank (n = 97,131 women) using the Global screening array genomic data (n = 535,107 SNPs). Univariate analyses revealed statistically

significant heritability (±SE) for age at menarche (43.1±5.8%, *n* = 8,071 women) and age at first birth (24.2±7.4%, *n* = 6,260 women), suggesting that genetic variance plays a significant role in shaping these traits ($p < 0.05$), but the heritability of the number of children was estimated with less precision and was not statistically significant (12.8±9.8%, $p = 0.337$, *n* = 4,602 women, Table 1a). This was also true for ER+ BC risk, where the heritability was estimated as 11.8% (±5.2%, $p = 0.190$, *n* = 182 cases and *n* = 8,700 controls, Table 1a). These heritability estimates were largely replicated using other genomic data available on Lifelines (the Affymetrix, *n* = 441,001 SNPs, and CytoSNP genomic array data, *n* = 244,041 SNPs, 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 (*n* = 198 cases, *n* = 8,972 controls). Moreover, the heritability estimate for age at menarche was considerably lower (but remained statistically significant) when using the CytoSNP data (17.0% ±5.2%, $p < 0.001$, *n* = 6,155 women, 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 family-based 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**

Genetic correlations (±SE) between ER+ BC risk and life-history traits were all low and statistically non-significant in the Lifelines data ($p > 0.01\dot{6}$). Figure 2a, Table S3). For the Global screening array data these were 0.05±0.17 for age at menarche (*n* = 176 cases, *n* = 7,895 controls), 0.10±0.24 for age at first birth (*n* = 155 cases, *n* = 6,105 controls), and 0.10±0.35 for number of children (n = 146 cases, n = 4,456 controls, $p > 0.016$, Fig 2a, Table S2). Precision of these estimates was also low, but a power analysis estimated that we had 80% power to detect genetic correlations between ER+BC risk and the life-history trait stronger than ±0.29 for age at menarche, ±0.40 for age at first birth, and ±0.52 for the number of children (Figure S1). These null findings were replicated when using both the Affymetrix and CytoSNP data (Table S3), with slightly reduced power in the CytoSNP data (Figure S1), which was expected given its lower sample size.

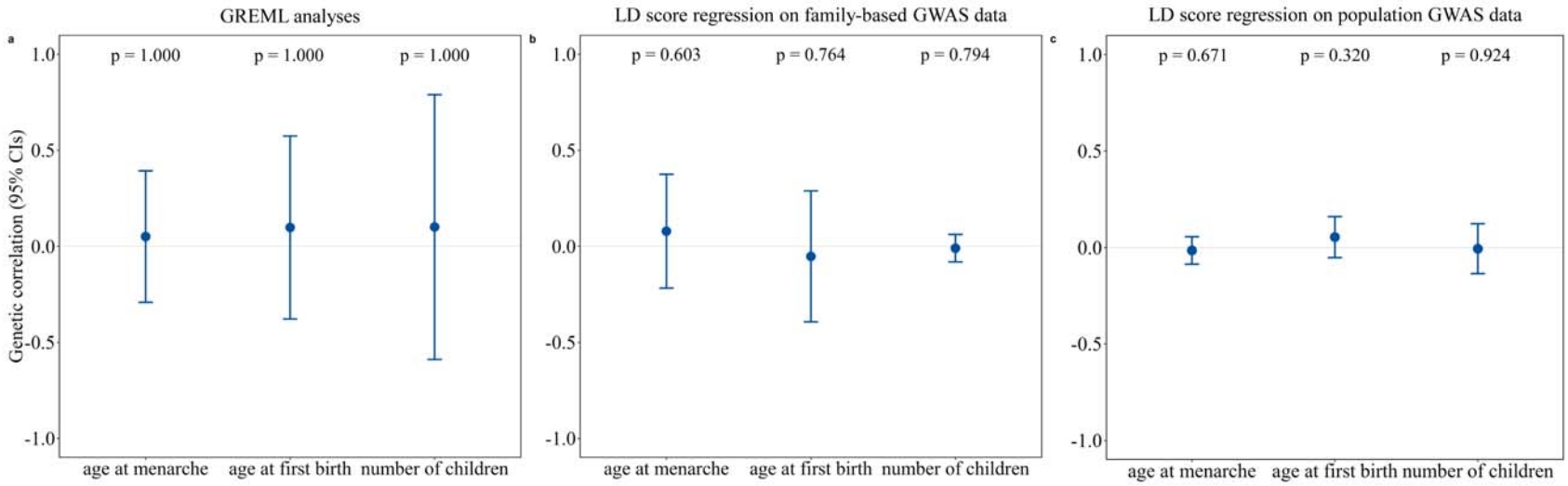


Figure 2: Genetic correlations between ER+ BC risk and life-history traits from: (**a**) a genomic relatedness matrix (GREML) analyses using Lifelines biobank data and LD score regression analyses using (**b**) family-based GWAS data and (**c**) population GWAS data. Circles above the dotted line indicate 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

To help us understand whether our results from the GREML analysis of Lifelines biobank data were limited by power or confounded by gene-environment correlations, we repeated our analysis 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 (sample sizes: n = 11,792,542 SNPs, n = 69,501 cases, and n = 95,042 controls, Michailidou et al., 2017) and the most recent family-based GWAS data study including life-history traits (n = 5,671,230-6,551,047 SNPs, Tan et al., 2024). Albeit estimated with more precision, our heritability estimates (±SE) 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%, n = 1,129,448 SNPs) and the number of children (8.6%±1.7%, n = 670,251 SNPs), 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% using the Global screening array data, n = 580,219 SNPs), although it remained statistically significant, whereas age at first birth was not shaped by genetic variation when using GWAS data (1.6%±1.9, n = 708,027 SNPs, 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 or the presence of assortative mating. A supplementary analysis of ER+ BC risk GWAS data also suggested little evidence of bias due to environmental confounding, including factors such as population stratification (Supplementary note 1, Table S5).

Using family-based GWAS summary statistics repeated the findings of no genetic correlations between life-history traits and ER+ BC risk, though our estimates were smaller and more precise (Figure 2b, Table S4, $p > 0.01\dot{6}$). Specifically, we estimated the genetic correlations (±SE) between ER+ BC risk and our life-history traits as: age at menarche, 0.079 (±0.15, $p = 0.602$, n = 551,452 SNPs); age at first birth, -0.052 (±0.178, $p = 0.764$, n = 668,389 SNPs); and the number of children ever born, -0.010 (±0.037, $p = 0.795$, n = 635,154 SNPs). 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.

Finally, we estimated genetic correlations using the population GWAS summary statistics for life-history traits uncorrected for population stratification from Tan et al. (2024) as family-based GWAS data can suffer from lower power and may underestimate genetic effects in the presence of assortative mating (Kemper et al., 2021). Using population GWAS data, we again repeated the findings of no statistically significant genetic correlations ($p > 0.01\dot{6}$), but estimates were generally estimated with more precision and, in some cases, varied quantitatively (Figure 2c, Table S6). Specifically, the genetic

correlation (±SE) between the life-history traits and ER+ BC risk was estimated as: for age at menarche, -0.015 (±0.036, p = 0.67, *n* = 886,799 SNPs), for age at first birth, 0.054 (±0.054, p = 0.32, *n* = 922,056 SNPs), and for number of children, -0.0063 (±0.066, p = 0.92, *n* = 952,783 SNPs, Table S6).

DISCUSSION

We estimated genetic correlations between three important human reproductive life-history traits (age at menarche, age at first birth, and number of children) and ER+ BC risk to help understand the extent to which life-history trade-offs shape genetic predispositions to ER+ BC. Our study builds on research showing evidence for antagonistic pleiotropy between ER+ BC risk and life-history traits at important loci underlying these traits (Day et al., 2017; Smith et al., 2012; Xiang et al., 2025; Zhao et al., 2024) to include the full contribution of pleiotropy to the covariance between these traits. Across all three life-history traits and all analyses we find no significant genetic correlations with ER+ BC risk. These results suggest that overall the importance of reproductive life-history trade-offs in shaping genetic ER+ BC risk is low and that genetic variation does not contribute overall to the observed phenotypic trends from epidemiological studies (Aktipis et al., 2015; Kelsey et al., 1993; Momenimovahed & Salehiniya, 2019). These results are in line with previous non-significant estimates from two studies (Jiang et al., 2019; Prince et al., 2024), with our analyses suggesting that these patterns hold for contemporary European ancestry populations regardless of the data or approach used, even when accounting for environmental confounding using family-based GWAS data.

Our study adds to previous research where studies focusing only on genome-wide significant loci show evidence for genetic life-history trade-offs shaping (ER+) BC risk (Day et al., 2017; Smith et al., 2012; Xiang et al., 2025; Zhao et al., 2024), whereas studies focusing on whole-genome genetic correlations find no evidence (Jiang et al., 2019; Prince et al., 2024). These results are not mutually exclusive, and our results do not refute the importance of life-history trade-offs being important in shaping individual ER+ BC risk variants. Especially in the case of high-prevalence loci of strong effect such as BRCA1/2 mutations (Smith et al., 2012) which are believed to be under strong negative selection in the absence of such pleiotropy (Pavard & Metcalf, 2007). However, because genetic correlations measure the average contribution of pleiotropy to the covariance between two traits (van Rheenen et al., 2019), together these results suggest that loci showing strong antagonistic pleiotropy are not repeatedly found across all loci underlying these complex, highly polygenic traits. Thus, rather than supporting the antagonistic pleiotropy theory of ageing, our results suggest that, in general, genetic ER+ BC risk variants may have accumulated due to weak selection at older ages, providing indirect support for the mutation accumulation theory of ageing (Charlesworth, 1980, 2001; Haldane, 1941; Medawar, 1952). Antagonistic pleiotropy between diseases acting in early- versus late-life can

also maintain genetic variation in disease risk, also in line with the antagonistic pleiotropy disease of ageing (Rodríguez et al., 2017). Thus, while our results suggest a minor role for reproductive life-history trade-offs in maintaining genetic variation in ER+ BC risk, studies estimating genetic correlations between ER+ BC risk and earlier onset diseases could help resolve the relative importance of the antagonistic pleiotropy or mutation accumulation theories of ageing for shaping genetic variation in this disease.

It is unclear whether these low and non-significant genetic correlations are due to: (1) little overall overlap between the genetic architecture of ER+ BC and these human life-history traits; and/or (2) variation in the direction of the associations across individual loci important for these sets of traits (van Rheenen et al., 2019). The first explanation is viable considering the polygenicity of these traits and was previously suggested as an explanation by Jiang et al. (2019). However, the current methods and data available cannot rule out the role of important breast cancer risk genes showing either agonistic or antagonistic, pleiotropy with life-history traits. With increasing power to detect genome-wide significant variants, future studies may be able to help assess the relative numbers of antagonistically versus agonistically pleiotropic variants and the degree of shared genetic architecture between ER+ BC risk and life-history traits.

Our results also provide no evidence for strong, consistent directional selection across ER+ BC risk loci, suggesting that these variants will not significantly increase or decrease in prevalence in future generations due to selection. This is despite demonstrating that contemporary European populations retain the ability to evolve through natural selection (Corbett et al., 2018; Fisher, 1930; Haldane, 1954). 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 trends would suggest that ER+ BC risk variants are under negative selection (Aktipis et al., 2015), our results suggest that these phenotypic trends do not translate to the genetic level (Haldane, 1954; Lande & Arnold, 1983; Robertson, 1966). Again, this does not rule out selection on individual loci, and this picture may change with additional, higher-power, GWAS studies. However, for traits such as the number of children, where the number of loci of large effect are relatively few (Barban et al., 2016), studies examining local genetic correlations, for example using ρHESS (Shi et al., 2017), would be best placed to detect finer-scale variation in the direction and strength of selection on genomic ER+ BC risk and assess the importance of selection and reproduction-cancer suppression trade-offs in shaping genetic ER+ BC risk. Additionally, recent research suggests that genetic variation in breast cancer is enriched in gene-expression hubs involved in estrogen metabolism and signaling (Stone et al., 2025). Given that these hubs are also likely to be important in regulating these life-history traits (Jacqueline et al., 2017; Jasienska et al., 2017), incorporating gene-networks and gene-expression data could further our understanding of how

variation at the genetic level could map to phenotypic trends, with studies also showing that transcriptomic correlations are generally higher than genetic correlations (Gerring et al., 2022).

No (or at most very weak) genetic correlations between life-history traits and ER+ BC risk also suggest that the observed phenotypic trends from epidemiological studies are largely due to non-genetic factors (Aktipis et al., 2015; Kelsey et al., 1993; Momenimovahed & Salehiniya, 2019). This provides indirect support for 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 phenotypic associations are environmental in origin. However, the extent to which changes in these reproductive life-history traits have also driven the increased prevalence in ER+ BC risk is unclear, which longitudinal studies would be best placed to answer.

Our study suffers from a few key limitations. First, although we found little evidence in support of life-history trade-offs for ER+ BC risk, exploring genetic correlations between estrogen receptor-negative breast cancer (ER- BC) risk and life-history traits may find different results. While ER- BC risk is not as strongly linked to reproductive life-history traits as ER+ BC risk at the phenotypic level (Aktipis et al., 2015), ER- BC risk does show higher heritability than ER+ BC (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). Second, we could not account for age at menopause in our analysis due to insufficient sample sizes. 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). Moreover, our study focused on reproductive life-history trade-offs, but body size – due to its impact on an organism's number of cells – is another important life-history trait affecting overall cancer susceptibility, including potentially ER+ BC risk (Aktipis et al., 2013; Nunney, 2018, 2024). Height has been identified as a potential risk factor for breast cancer (Green et al., 2011), although current evidence of this being due to genetic factors is lacking (Jiang et al., 2019), but warrants further study. Finally, despite using both individual-level biobank data and the largest available GWAS data for life-history traits and ER+ BC risk, some of our genetic correlations were still somewhat imprecise. These

power limitations were mostly constrained to the GREML analyses of Lifelines biobank data, where some traits in some analyses did not show significant heritability (e.g., ER+ BC risk and number of children when using the Global screening array and CytoSNP array data). However, even when some of these traits showed higher and statistically significant heritability (e.g., using the Affymetrix data), the finding of non-significant genetic correlations was replicated, although we could only rule out genetic correlations larger than ±0.29-0.52 for the Global screening array data. Estimating genetic correlations with GREML is computationally demanding but further studies could aim to conduct similar analyses using larger biobanks. Nevertheless, our analyses of GWAS data were less power-limited and more precise. Our study therefore provides a robust assessment of the strength of genetic correlations between ER+ BC risk and these human life-history traits, given the current GWAS data and methods available.

These findings provide an important step towards a more complete understanding of the genetic and evolutionary processes shaping ER+ BC risk, which is a deadly form of cancer with increasing prevalence. Our results suggest that genetic predispositions towards ER+ BC risk are not overall shaped by life-history trade-offs in European ancestry populations but whether this is also true across other geographic regions is unknown. Studies of populations with higher birth rates, increasing the chance of detecting reproduction-survival trade-offs, have previously been used to identify antagonistically pleiotropic genes underlying disease and fertility (e.g., Trumble et al., 2023), could be particularly fruitful. Moreover, 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 S2A), 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 S2B) 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 of childless women and the total number of children born. This gave a mean number of children born of 2.0 (range 0-8, N = 51,672, Figure S2C). All data were handled in R 4.2.2 (2022-10-31, R Core Team, 2020) using *dplyr* 1.1.4 (Wickham et al., 2019).

ER+ BC diagnoses

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

We subtyped these records according to estrogen receptor status (ER-status) using the report conclusions written by pathologists contained in the Palga excerpts (see Supplementary note 2 for further details). According to national Dutch guidelines, ER-status is considered positive when >= 10% of tumor cells show are ER+ (IKNL, 2012). We focused on women whose first cancer diagnosis was breast cancer to remove possible metastatic tumors of different origin (n = 628). We also excluded 15 male diagnoses. This left 3,667 Lifelines participants with their first-cancer diagnosis being breast cancer, diagnosed from 1981-2023, and 2,177 ER+ diagnoses (Figure S3), 415 ER- diagnoses, with 1,090 being unable to be assigned an ER-status. The ER+ BC diagnoses had a median age of diagnosis of 51 (range = 24-93),

Genotype data

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

We aimed to combine the different Lifelines genotypes in one GREML analysis and therefore harmonized these three datasets using Genotype Harmonizer 1.4.25 (available at 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:

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

We created the genomic relatedness matrices (GRMs) for all individuals for which we had genomic data using the *--make-grm* argument in 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*-women = 25,068) to remove bias in genetic variances due to 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+ BC, was included as a linear and squared covariate to control for potential age effects while maximizing available sample sizes. The undiagnosed age could be the age of the individual at the date the Palga records were up to date until (February 21st, 2023) or the date of death for an individual if this occurred prior (n = 2,293). Dates of death were included in Lifelines as recorded by the Dutch government and were known even in cases of emigration. These dates were decimalized using *lubridate* 1.9.3 in R (Grolemund & Wickham, 2011) and ages were calculated using the birth year and month available in Lifelines. We also included the participant's birthdate as a linear and squared covariate, to account for any temporal trends in variables, and the top 20 principal components of the pruned GRM used in each model to account for possible population stratification (Yang et al., 2014). Principal components were generated using the *--pca* argument in the *gcta64* function.

However, our bivariate models struggled to converge and univariate models provided apparently biased heritability estimates inconsistent with the literature (e.g., 0% for age at menarche, Table S1d). We posited that this was due to poor coverage of loci between the different genomic samples. For example, although we had over 1 million SNPs in total, the number of SNPs measured in individuals belonging to, for example, both the Affymetrix and CytoSNP array samples was ~2% of the total number of SNPs available (Table S2). This could result in biased genomic relatedness estimates and increase the chance of missing causal loci, downwardly biasing heritability estimates in the univariate models (Yang et al., 2010). We therefore created GRMs based on each genomic dataset separately where the number of loci measured between individuals was larger and more consistent (Table S2). We present the analysis using the Global screening array data in the main text – because it was the largest in both number of SNPs and individuals – and use the analyses of the other cohorts as replicates. Our sample sizes of informative individuals for our analyses using the Global screening array were: age at menarche, 8,071; age at first birth, 6,260; number of children ever born, 4,602; and 182 ER+ BC cases and 8,700 controls. We also performed a power analysis for our genetic correlations estimated from GREML analyses for each of our traits and genomic datasets based upon heritability estimates from univariate analyses and case and control sample sizes (Figure S1, following Visscher et al., 2014)

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

LD score regression of GWAS data

We then replicated our findings from the genomic relatedness analyses using LD score regression to estimate heritabilities and genetic correlations between our traits from previously published GWAS summary 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; A. I. 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 $E\left[\chi_j^2\right]$ 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 $z_{1j} z_2 l_j$ for a SNP *j* is

$$E\left[z_{1j}z_{2}l_{j}\right] = \frac{\sqrt{N_1N_2}\varrho_g}{M}l_j + \frac{\varrho N_s}{\sqrt{N_1N_2}}$$

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:

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

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). Code for replicating data wrangling and analysis is available for reviewers here: https://dataverse.nl/privateurl.xhtml?token=93828d52-1a89-4e46-8244-7e21eaf987cf.

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. V.L. was funded by the Strategic Research Council of the

Academy of Finland (grant nos. 345185 and 345183) and an European Research Grant (ERC-2022-ADG 101098266).

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, Raphaël Scherrer, Judith Vonk, Harold Snieder, and Ilja Nolte for their advice on the data handling and analysis. 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. We thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Hábrók high performance computing cluster.

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., Gatenby, R. A., Brown, J. S., & Maley, C. C. (2013). Life history trade-offs in cancer evolution. *Nature Reviews Cancer*, *13*(12), 883–892. https://doi.org/10.1038/nrc3606

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

Bauer, K. R., Brown, M., Cress, R. D., Parise, C. A., & Caggiano, V. (2007). Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: A population-based study from the California cancer Registry. *Cancer*, *109*(9), 1721–1728. https://doi.org/10.1002/cncr.22618

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

Boddy, A. M., Kokko, H., Breden, F., Wilkinson, G. S., & Aktipis, C. A. (2015). Cancer susceptibility and reproductive trade-offs: A model of the evolution of cancer defences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1673), 20140220. https://doi.org/10.1098/rstb.2014.0220

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 Wellcome Trust Case Control Consortium 3, Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Consortium, S. W. G. of the P. G., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211

Byars, S. G., Huang, Q. Q., Gray, L.-A., Bakshi, A., Ripatti, S., Abraham, G., Stearns, S. C., & Inouye, M. (2017). Genetic loci associated with coronary artery disease harbor evidence of selection and antagonistic pleiotropy. *PLOS Genetics*, *13*(6), e1006328. https://doi.org/10.1371/journal.pgen.1006328

Byars, S. G., & Voskarides, K. (2020). Antagonistic Pleiotropy in Human Disease. *Journal of Molecular Evolution*, *88*(1), 12–25. https://doi.org/10.1007/s00239-019-09923-2

Casparie, M., Tiebosch, A. T. M. G., Burger, G., Blauwgeers, H., Van De Pol, A., Van Krieken, J. H. J. M., & Meijer, G. A. (2007). Pathology Databanking and Biobanking in The Netherlands, a Central Role for PALGA, the Nationwide Histopathology and Cytopathology Data Network and Archive. *Analytical Cellular Pathology*, *29*(1), 19–24. https://doi.org/10.1155/2007/971816

Charlesworth, B. (1980). *Evolution in age-structured populations*. Cambridge University Press Cambridge.

Charlesworth, B. (2001). Patterns of age-specific means and genetic variances of mortality rates predicted by the mutation-accumulation theory of ageing. *Journal of Theoretical Biology*, *210*(1), 47–65.

Coe, K., & Steadman, L. B. (1995). The human breast and the ancestral reproductive cycle: A preliminary inquiry into breast cancer etiology. *Human Nature*, *6*(3), 197–220. https://doi.org/10.1007/BF02734139

Colditz, G. A., Rosner, B. A., Chen, W. Y., Holmes, M. D., & Hankinson, S. E. (2004). Risk Factors for Breast Cancer According to Estrogen and Progesterone Receptor Status. *JNCI Journal of the National Cancer Institute*, *96*(3), 218–228. https://doi.org/10.1093/jnci/djh025

Corbett, S., Courtiol, A., Lummaa, V., Moorad, J., & Stearns, S. (2018). The transition to modernity and chronic disease: Mismatch and natural selection. *Nature Reviews Genetics*, *19*(7), 419–430. https://doi.org/10.1038/s41576-018-0012-3

Crespi, B. J., & Summers, K. (2006). Positive selection in the evolution of cancer. *Biological Reviews of the Cambridge Philosophical Society*, *81*(3), 407–424. https://doi.org/10.1017/S1464793106007056

Dall, G. V., Hawthorne, S., Seyed-Razavi, Y., Vieusseux, J., Wu, W., Gustafsson, J.-A., Byrne, D., Murphy, L., Risbridger, G. P., & Britt, K. L. (2018). Estrogen receptor subtypes dictate the proliferative nature of the mammary gland. *Journal of Endocrinology*, *237*(3), 323–336. 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* [Data set]. Global Burden of Disease Collaborative Network, Institute for Health Metrics and Evaluation (IHME). https://vizhub.healthdata.org/gbd-results/

Gerring, Z. F., Thorp, J. G., Gamazon, E. R., & Derks, E. M. (2022). An analysis of genetically regulated gene expression and the role of co-expression networks across 16 psychiatric and substance use phenotypes. *European Journal of Human Genetics*, *30*(5), 560–566. https://doi.org/10.1038/s41431-022-01037-6

Ghosh, M. G., Thompson, D. A., & Weigel, R. J. (2020). PDZK1 and GREB1 Are Estrogen-regulated Genes Expressed in Hormone-responsive Breast Cancer. *Cancer Research*, *60*, 6367–6375.

Green, J., Cairns, B. J., Casabonne, D., Wright, F. L., Reeves, G., & Beral, V. (2011). Height and cancer incidence in the Million Women Study: Prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *The Lancet Oncology*, *12*(8), 785–794. https://doi.org/10.1016/S1470-2045(11)70154-1

Grolemund, G., & Wickham, H. (2011). Dates and Times Made Easy with {lubridate}. *Journal of Statistical Software*, *40*(3), 1–25. https://www.jstatsoft.org/v40/i03/

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

Jasienska, G., Bribiescas, R. G., Furberg, A. S., Helle, S., & Núñez-de la Mora, A. (2017). Human reproduction and health: An evolutionary perspective. *The Lancet*, *390*(10093), 510–520. https://doi.org/10.1016/S0140-6736(17)30573-1

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

Kang, H. J., Feng, Z., Sun, Y., Atwal, G., Murphy, M. E., Rebbeck, T. R., Rosenwaks, Z., Levine, A. J., & Hu, W. (2009). Single-nucleotide polymorphisms in the p53 pathway regulate fertility in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(24), 9761–9766. https://doi.org/10.1073/pnas.0904280106

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.

Kelsey, J. L., Gammon, M. D., & John, E. M. (1993). Reproductive factors and breast cancer. *Epidemiologic Reviews*, *15*(1), 36–47. https://doi.org/10.1093/oxfordjournals.epirev.a036115

Kemper, K. E., Yengo, L., Zheng, Z., Abdellaoui, A., Keller, M. C., Goddard, M. E., Wray, N. R., Yang, J., & Visscher, P. M. (2021). Phenotypic covariance across the entire spectrum of relatedness for 86 billion pairs of individuals. *Nature Communications*, *12*(1), 1050. https://doi.org/10.1038/s41467-021-21283-4

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). [Data set]. https://data-catalogue.lifelines.nl/

Long, E., & Zhang, J. (2023). Evidence for the role of selection for reproductively advantageous alleles in human aging. *Science Adanvces*, *9*(eadh4990), 1–11.

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., Keeman, R., Eilber, U., … Easton, D. F. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, *551*(7678), 92–94. https://doi.org/10.1038/nature24284

Michailidou, K., The Breast and Ovarian Cancer Susceptibility Collaboration, Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON), kConFab Investigators, Australian Ovarian Cancer Study Group, The GENICA (Gene Environment Interaction and Breast Cancer in Germany) Network, Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L., Schmidt, M. K., Chang-Claude, J., Bojesen, S. E., Bolla, M. K., Wang, Q., Dicks, E., Lee, A., Turnbull, C., … Easton, D. F. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature Genetics*, *45*(4), 353–361. https://doi.org/10.1038/ng.2563

Mills, M. C., Tropf, F. C., Brazel, D. M., van Zuydam, N., Vaez, A., Agbessi, M., Ahsan, H., Alves, I., Andiappan, A. K., Arindrarto, W., Awadalla, P., Battle, A., Beutner, F., Jan Bonder, M., Boomsma, D. I., Christiansen, M. W., Claringbould, A., Deelen, P., Esko, T., … Day, F. R. (2021). Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01135-3

Momenimovahed, Z., & Salehiniya, H. (2019). Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast Cancer: Targets and Therapy*, *Volume 11*, 151–164. https://doi.org/10.2147/BCTT.S176070

Nguyen, N., Fan, H.-Y., Tsai, M.-C., Tung, T.-H., Huynh, Q., Huang, S.-Y., & Chen, Y. (2020). Nutrient Intake through Childhood and Early Menarche Onset in Girls: Systematic Review and Meta-Analysis. *Nutrients*, *12*(9), 2544. https://doi.org/10.3390/nu12092544

Nunney, L. (2013). The real war on cancer: The evolutionary dynamics of cancer suppression. *Evolutionary Applications*, *6*(1), 11–19. https://doi.org/10.1111/eva.12018

Nunney, L. (2018). Size matters: Height, cell number and a person's risk of cancer. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1889), 20181743. https://doi.org/10.1098/rspb.2018.1743

Nunney, L. (2024). The effect of body size and inbreeding on cancer mortality in breeds of the domestic dog: A test of the multi-stage model of carcinogenesis. *Royal Society Open Science*, *11*(1), 231356. https://doi.org/10.1098/rsos.231356

Parent, A.-S., Teilmann, G., Juul, A., Skakkebaek, N. E., Toppari, J., & Bourguignon, J.-P. (2003). The Timing of Normal Puberty and the Age Limits of Sexual Precocity: Variations around the World, Secular Trends, and Changes after Migration. *Endocrine Reviews*, *24*(5), 668–693. https://doi.org/10.1210/er.2002-0019

Pavard, S., & Metcalf, C. J. E. (2007). Negative Selection on BRCA1 Susceptibility Alleles Sheds Light on the Population Genetics of Late-Onset Diseases and Aging Theory. *PLoS ONE*, *2*(11), e1206. https://doi.org/10.1371/journal.pone.0001206

Perry, J. R. B., Murray, A., Day, F. R., & Ong, K. K. (2015). Molecular insights into the aetiology of female reproductive ageing. *Nature Reviews Endocrinology*, *11*(12), 725–734. 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

R Core Team. (2020). *R: A language environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. http://www.r-project.org/.

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

Shi, H., Mancuso, N., Spendlove, S., & Pasaniuc, B. (2017). Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. *The American Journal of Human Genetics*, *101*(5), 737–751. https://doi.org/10.1016/j.ajhg.2017.09.022

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; WorldCat.

Stone, K. L., Platig, J., Quackenbush, J., & Fagny, M. (2025). The Importance of Regulatory Network Structure for Complex Trait Heritability and Evolution. *Molecular Biology and Evolution*, *42*(8), msaf174. https://doi.org/10.1093/molbev/msaf174

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

Summers, K., & Crespi, B. (2008). The androgen receptor and prostate cancer: A role for sexual selection and sexual conflict? *Medical Hypotheses*, *70*(2), 435–443. https://doi.org/10.1016/j.mehy.2007.04.044

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

Tropf, F. C., Stulp, G., Barban, N., Visscher, P. M., Yang, J., Snieder, H., & Mills, M. C. (2015). Human fertility, molecular genetics, and natural selection in modern societies. *PLoS ONE*, *10*(6), 1–14. https://doi.org/10.1371/journal.pone.0126821

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*.

van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. (2019). Genetic correlations of polygenic disease traits: From theory to practice. *Nature Reviews Genetics*, *20*(10), 567–581. https://doi.org/10.1038/s41576-019-0137-z

Visscher, P. M., Hemani, G., Vinkhuyzen, A. A. E., Chen, G.-B., Lee, S. H., Wray, N. R., Goddard, M. E., & Yang, J. (2014). Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. *PLoS Genetics*, *10*(4), e1004269. https://doi.org/10.1371/journal.pgen.1004269

Wang, X., Byars, S., & Stearns, S. (2013). Genetic links between post-reproductive lifespan and family size in Framingham. *Evolution Medicine & Public Health*, *2013*(1), 241–253. https://doi.org/10.1093/emph/eot013

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

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

Xiang, Y., Tanwar, V., Singh, P., La Follette, L., Narayan, V. P., & Kapahi, P. (2025). Early menarche and childbirth accelerate aging-related outcomes and age-related diseases: Evidence for antagonistic pleiotropy in humans. *eLife*, *13*, RP102447. https://doi.org/10.7554/eLife.102447

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.

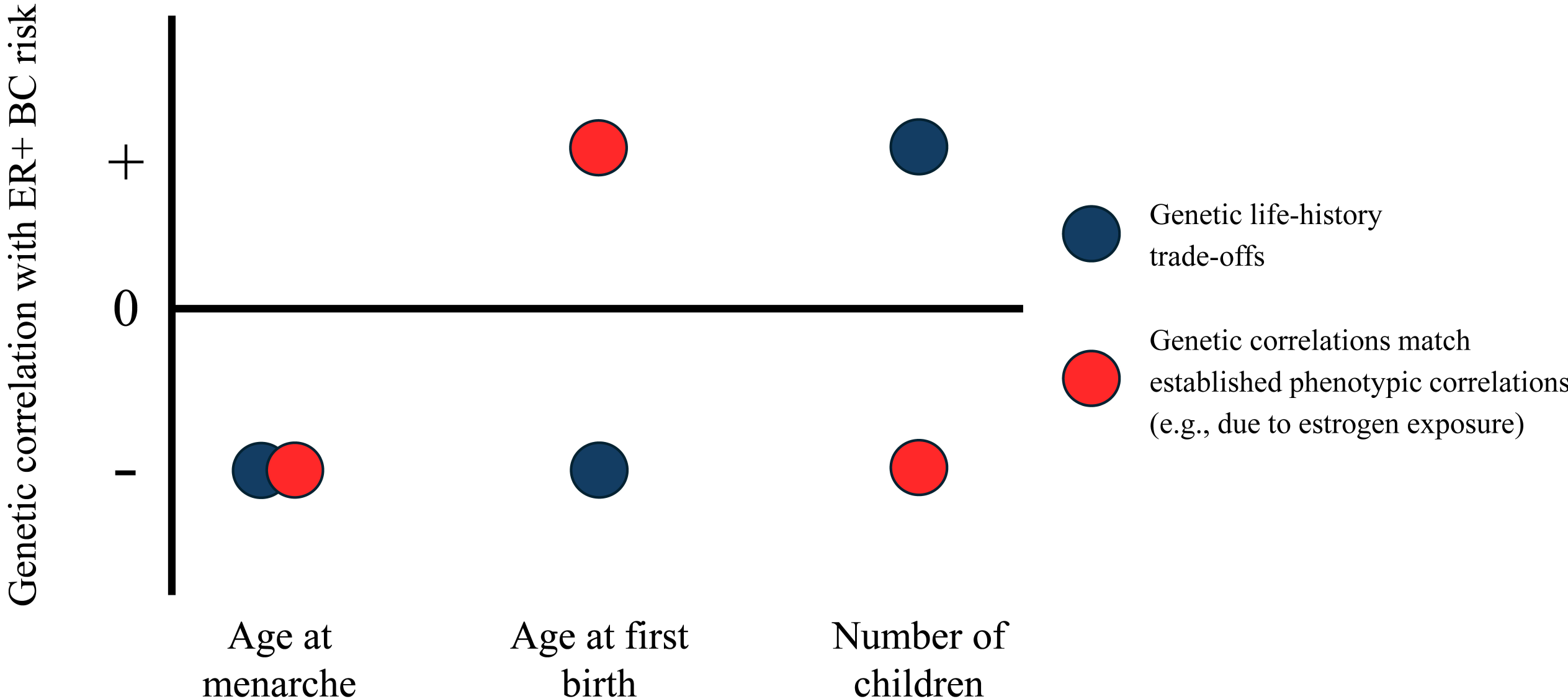
You, W., Symonds, I., Rühli, F. J., & Henneberg, M. (2018). Decreasing Birth Rate Determining Worldwide Incidence and Regional Variation of Female Breast Cancer. *Advances in Breast Cancer Research*, *07*(01), 1–14. https://doi.org/10.4236/abcr.2018.71001

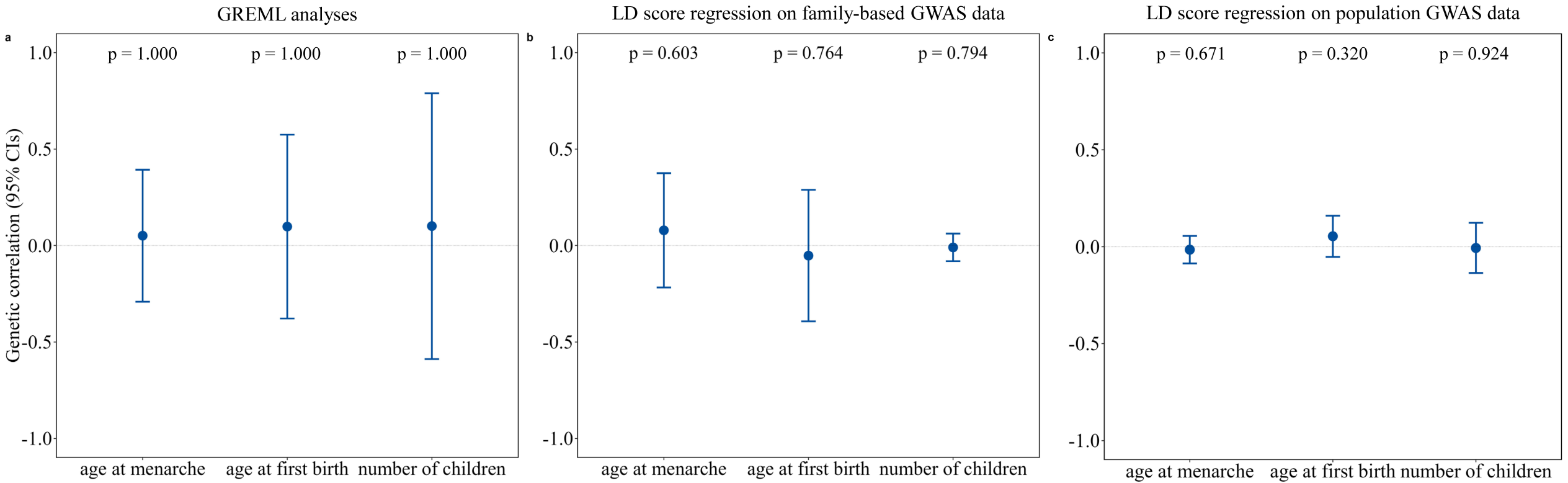
Young, A. I., Nehzati, S. M., Benonisdottir, S., Okbay, A., Jayashankar, H., Lee, C., Cesarini, D., Benjamin, D. J., Turley, P., & Kong, A. (2022). Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nature Genetics*, *54*(6), 897–905. https://doi.org/10.1038/s41588-022-01085-0

Young, A. S. (2024). Genome-wide association studies have problems due to confounding: Are family-based designs the answer? *PLOS Biology*, *22*(4), e3002568. https://doi.org/10.1371/journal.pbio.3002568

Zhang, H., Ahearn, T. U., Lecarpentier, J., Barnes, D., Beesley, J., Qi, G., Jiang, X., O'Mara, T. A., Zhao, N., Bolla, M. K., Dunning, A. M., Dennis, J., Wang, Q., Ful, Z. A., Aittomäki, K., Andrulis, I. L., Anton-Culver, H., Arndt, V., Aronson, K. J., … García-Closas, M. (2020). Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nature Genetics*, *52*(6), 572–581. https://doi.org/10.1038/ng.1049

Zhao, Z., Zhang, J., & Tian, X. (2024). Relationship between age at menarche and breast cancer in individuals, as well as in first-degree kin and estrogen receptor status: A Mendelian randomization study. *Frontiers in Oncology*, *14*, 1408132. https://doi.org/10.3389/fonc.2024.1408132


Genetic correlation with ER+ BC risk
+
0
-
Age at menarche
Age at first birth
Number of children
Genetic life-history trade-offs
Genetic correlations match established phenotypic correlations (e.g., due to estrogen exposure)


GREML analyses
LD score regression on family-based GWAS data
LD score regression on population GWAS data
a
b
c
Genetic correlation (95% CIs)
1.0
0.5
0.0
-0.5
-1.0
p = 1.000
p = 1.000
p = 1.000
p = 0.603
p = 0.764
p = 0.794
p = 0.671
p = 0.320
p = 0.924
age at menarche
age at first birth
number of children