Article

Increasing number of long-lived ancestors marks a decade of healthspan extension and healthier metabolomics profiles

Niels van den Berg [1,2] ✉, Mar Rodríguez-Girondo[3], Ingrid K. van Dijk[2], P. Eline Slagboom [1,4,5] & Marian Beekman [1,5]

Globally, the lifespan of populations increases but the healthspan is lagging behind. Previous research showed that survival into extreme ages (longevity) clusters in families as illustrated by the increasing lifespan of study participants with each additional long-lived family member. Here we investigate whether the healthspan in such families follows a similar quantitative pattern using three-generational data from two databases, LLS (Netherlands), and SEDD (Sweden). We study healthspan in 2143 families containing index persons with 26 follow-up years and two ancestral generations, comprising 17,539 persons. Our results provide strong evidence that an increasing number of long-lived ancestors associates with up to a decade of healthspan extension. Further evidence indicates that members of long-lived families have a delayed onset of medication use, multimorbidity and, in mid-life, healthier metabolomic profiles than their partners. We conclude that both lifespan and healthspan are quantitatively linked to ancestral longevity, making family data invaluable to identify protective mechanisms of multimorbidity.

The human life expectancy has doubled over the past two centuries[1], reaching 82.1 years in Western European countries[2]. Although people started to live longer, the time spent in good physical and cognitive health did not rise similarly[2]. In fact, over 70% of the 65 year olds have at least one disease and over 50% have multimorbidity (2 disease or more)[3]. In contrast to the general population, some persons seem to become exceptionally old with a significantly lower chronic age-related disease burden (e.g. high blood pressure, malignancies, and type 2 diabetes) than the general population[4–17]. Moreover, the children of these exceptionally old persons have a delayed first disease onset[11,14,18,19]. These observations are mostly based on cross-sectional designs[4,9–14,17–19], so prospective studies into the development of first diseases and (multi)morbidity are needed. The study of long-lived families is important as they likely harbor gene-environment

interactions which beneficially regulate molecular pathways involved in longevity, resistance to disease, resilience to negative side-effects of treatment, and therefore healthy aging[8,20].

In our previous work, we used data from millions of subjects in contemporary medical and historical family-tree databases to investigate the intergenerational transmission of human longevity[21–24]. We concluded that longevity, as a heritable trait, is primarily transmitted if persons belong to the oldest 10% survivors of their birth cohort and if at least 30% of their ancestors also belonged to the oldest 10% survivors[21,22]. Subsequently, we developed the Longevity Relatives Count (LRC) score as an instrument to quantify the number of long-lived family members and observed that the survival advantage of study participants increased with each additional long-lived family member, indicating additive effects[21]. As such, the LRC score is an

[1]Department of Biomedical Data Sciences, section of Molecular Epidemiology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands. [2]Centre for Economic Demography, Department of Economic History, Lund University, Scheelevägen 15B, 223 63 Lund, Sweden. [3]Department of Biomedical Data Sciences, section of Medical Statistics, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands. [4]Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Str. 9b, D-50931 Cologne, Germany. [5]These authors contributed equally: P. Eline Slagboom, Marian Beekman. ✉e-mail: n.m.a.van_den_berg@lumc.nl

indicator of increased survival and longevity, and can therefore be used to enlarge the survival contrast in epidemiological data, thereby leading to more powerful genetic longevity studies. If the LRC score also represents healthspan as a quantitative trait (additive effects), this instrument can potentially be used in (genetic) studies to elucidate multimorbidity-limiting mechanisms.

We identified two issues that have not yet thoroughly been investigated: (1) whether from mid-life onward, health, medication use, disease incidence as well as the development of multimorbidity are delayed over time, and (2) whether an increasing number of long-lived ancestors, as measured with the LRC score, represents not only lifespan as a quantitative trait but also healthspan. To address these issues, longitudinal life course and health data should ideally be investigated, preferably in large numbers of individuals. In addition, multigenerational family-tree information is required to investigate how the number of ancestral long-lived relatives relates to morbidity. Therefore we investigate chronic disease incidence and multi-morbidity in long-lived families using up to 25 years of follow-up data. We further study whether an increasing number of long-lived ancestors, as measured by the LRC score, associates with a decreased incidence of chronic diseases. In addition, we investigate whether the families with the most extreme LRC scores have a healthy metabolomic profile in mid-life, representing overall health to complement the information on morbidity.

We use the data available in the Leiden Longevity Study (LLS, Netherlands) and the Swedish register data available in the Scanian Economic-Demographic Database (SEDD, Sweden). The LLS, initiated in 2002, was based on the inclusion of nonagenarian siblings. In addition, the middle-aged children (called index persons (IPs) in the current study) and their partners, as adult environment-matched controls, were included. The SEDD contains the entire population of 5 parishes and a town in Scania (southern Sweden), and as such does not contain any initial inclusion criteria. For the current study we identified in both datasets combined 2143 three-generational families (F1-F3) containing IPs (F3) and their family members, comprising 17,539 persons in total. First, we examine whether LLS IPs and their partners differ in terms of disease and medication prevalence at the moment of study inclusion (2002–2004). Second, we investigate differences in disease incidence towards multimorbidity. Third, we study whether an increasing number of long-lived ancestors is associated with a decreased disease incidence in IPs using the Longevity Relatives Count (LRC) score[22] in both the LLS and SEDD datasets. Finally, we compare mid-life health of LLS IPs with the highest LRC scores and their partners, using a previously developed metabolomics-based score predicting mortality[25].

Results
Study populations
LLS IPs and their partners, serving as environment-matched controls, were included between 2002 and 2006 at the average age of 59 years. The study inclusion was based on nonagenarian siblings in the F2 generation. Hence, IPs (F3) were included if they had at least one long-lived F2 parent and F2 aunt or uncle (females ≥91 years and males ≥89 years). The LLS consists of 651 three-generational families, defined by IP siblings who have the same parents (mean sibship size is 2.58). From inclusion onward, the IPs and their partners were followed over time, with a maximum mortality follow-up of 19 years (2002–2021) and maximum morbidity follow-up of 16 years (2002–2018). In 2021, 227 (14%) IPs and 113 (15%) partners were deceased and 1409 (84%) IPs and 619 (83%) partners were still alive. In 2018, 671 (40%) IPs and 324 (43%) partners had a disease diagnosis whereas 535 (32%) IPs and 206 (28%) partners did not have a disease diagnosis (Fig. 1a and Table 1A).

The SEDD consists of 1495 three-generational families, defined by IP siblings who have the same parents (mean sibship size is 1.67). SEDD

IPs were followed over time from 1990, at an average age of 52 years, with a maximum mortality and morbidity follow-up of 26 years (1990–2015). In 2015, 694 (28%) IPs were deceased whereas 1803 (72%) were still alive. Moreover, 1190 (48%) IPs had a disease diagnosis whereas 1307 (52%) IPs did not have a disease diagnosis (Fig. 1B and Table 1B). From here we will refer to disease diagnoses as diseases, disease prevalence in cross-sectional analyses, and disease incidence in longitudinal analyses.

IPs have a lower risk of using medication early in the study
Data on medication use was collected in the LLS between 2006 and 2008 (Supplementary Fig. 1) and was available for 1254 LLS IPs (75%) and 588 partners (79%). We focused on ATC A-C type medication because they match the disease groups we investigate (see "Methods" section and Supplementary Table 4). To study whether LLS IPs had a lower medication use compared to their partners, we fitted mixed-model logistic regression analyses. Figure 2 shows that the odds of using ATC-A (alimentary tract and metabolism) type medication is 45% (OR = 0.55 (95% CI = 0.42−0.71)) lower for the offspring than for their partners. Similarly, the odds of using ATC-B (blood and blood forming organs) and ATC-C (cardiovascular system) type medication is 42% (OR = 0.58 (95% CI = 0.37−0.92)) and 48% (OR = 0.52 (95% CI = 0.38−0.71)) lower for the IPs. Our analyses thus indicate that, early on in the study, the IPs already had a significantly lower prescription of metabolic and cardiovascular disease medication.
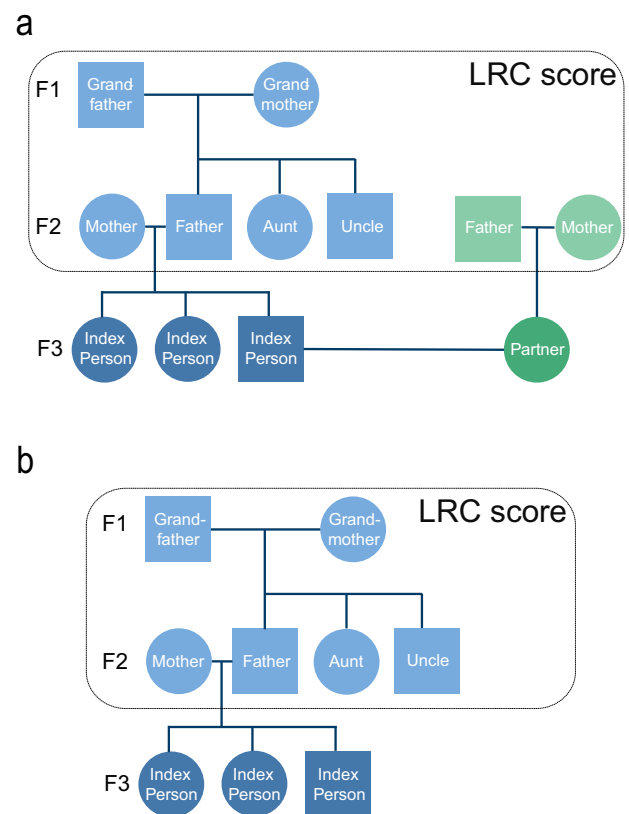
Fig. 1 | Conceptual pedigree of a 3 filial (F) generation LLS and SEDD family. This figure corresponds to Table 1. **a** represents a hypothetical family from the LLS and **b** represents a hypothetical family from the SEDD, both covering 3 filial (F) generations. Circles represent women, Squares represent men. Dark blue: index persons (IPs, F3), dark green: partners of IPs (F3; LLS only), light blue: fathers and mothers of IPs (F2), aunts and uncles of IPs (F2), grandmothers and grandfathers of IPs (F1), light green: fathers and mothers of IPs (F2). The F3 generation is investigated in this study and the F2 and F1 generations were used to calculate the Longevity Relatives Count (LRC) score. In the LLS, the mean sibship size is 2.58 and in SEDD the mean sibship size is 1.67.

Table 1 | Basic characteristics of LLS Index Persons, partners, and ancestral groups

	Analyses groups		Ancestral family groups			
	IPs (F3)	Partners of IPs (F3)	Parents of IPs (F2)	Aunts/uncles of IPs (F2)	Grandparents of IPs (F1)	Parents of partners (F2)
Panel A: LLS						
Number, *N* (number of families)	1674 (651)	745 (745)	1295	2370	760	1237
Female, *N* (%)	905 (54)	429 (57)	646 (50)	1169 (49)	380 (50)	620 (50)
Range birth cohorts, years	1923–1971	1924–1974	1882–1928	1875–1951	1850–1894	1864–1947
Alive, *N* (%)	1409 (84)	619 (83)	116 (9)	368 (16)	0 (0)	254 (21)
Deceased, *N* (%)	227 (14)	113 (15)	1179 (91)	2001 (84)	760 (100)	1011 (82)
Missing age, *N* (%)	38 (2)	13 (2)	94 (7)	34 (1)	0 (0)	35 (3)
Mean ad or al, years (SD)	74.87 (6.67)	74.21 (7.22)	87.09 (14.13)	71.93 (28.03)	77.06 (14.15)	76.68 (13.86)
Disease free, *N* (%)	535 (32)	206 (28)	–	–	–	–
Diseased, *N* (%)	671 (40)	324 (43)	–	–	–	–
Missing disease, *N* (%)	468 (28)	215 (29)	–	–	–	–
Mean number of diseases, *N*, (SD)	0.82 (0.92)	0.94 (0.98)	–	–	–	–
Panel B: SEDD						
Number, *N* (number of families)	2497 (1495)	–	2969	5830	3028	–
Female, *N* (%)	1252 (50.1)	–	1480 (49.8)	2866 (49.2)	1532 (50.6)	–
Range birth cohorts, years	1930–1950	–	1853–1914	1860-1950	1833-1918	–
Alive, *N* (%)	1803 (72.2)	–	7 (0.2)	855 (14.7)	131 (4.3)	–
Deceased, *N* (%)	694 (27.8)	–	2962 (99.8)	4975 (85.3)	2897 (95.7)	–
Missing age, *N* (%)	–	–	–	–	–	–
Mean ad or al, years (SD)	74.6 (6.7)	–	76.0 (17.9)	57.4 (32.2)	64.0 (19.6)	–
Disease free, *N* (%)	1307	–	–	–	–	–
Diseased, *N* (%)	1190	–	–	–	–	–
Missing disease, *N* (%)	–	–	–	–	–	–
Mean number of diseases, *N* (SD)	–	–	–	–	–	–

Index Persons (IPs) and their partners were included in the Leiden Longevity Study (LLS) between 2002 and 2006. The table is separated by analyses groups and ancestral family groups. The analyses group information represent the persons (IPs and their partners) that are investigated in this study, as a complete group or in different subgroups. The ancestral family information represents the ancestors of the IPs and their partners which are used to calculate the Longevity Relatives count (LRC) score. Panel A: For all groups, except the parents of partners (F2), mortality information was updated in January 2021. All mortality information was obtained from the official and verified Netherlands Population Registers. Mortality information for the parents of partners (F2) was obtained from questionnaires filled in by the partners of IPs (F3). Disease (morbidity) information was updated in 2018 based on input from the General Practitioners of the IPs and their partners. The study covers a maximal mortality follow-up of 19 years (2002–2021) and maximal morbidity follow-up of 16 years (2002–2018). Panel B: SEDD IPs were followed over time from 1990, at an average age of 52 years, with a maximal mortality and morbidity (disease) follow-up of 26 years (1990–2015). All mortality and morbidity data was obtained from the SEDD which is linked to the National Swedish Registers (see "Methods" for more details). Because register data was used, there was no missing data for ages and diseases. Moreover, we focused on first diseases and therefore no mean number of diseases can be provided for the SEDD data. Families are defined by siblings who have the same parents. ad = age at death and al = age at last observation.

IPs have a delayed first disease onset during follow-up

To investigate whether and to what extent the onset of first disease was delayed for the LLS IPs compared to their partners during 16 years of follow-up, we excluded persons who had ≥1 disease at inclusion. We therefore include 917 LLS IPs of whom 39 (4.3%) were deceased at the end of disease follow-up (2018) and 395 partners of whom 17 (4.3%) were deceased. We fitted random effect (frailty) Cox regressions and observed a Hazard Ratio (HR) of 0.79 (95% CI = 0.65–0.97) for the age-related disease incidence between LLS IPs and their partners. This HR indicates that the yearly risk of age-related disease was 21% lower for the LLS IPs as compared to their partners. The LLS IPs had a 29% (HR = 0.71 (95% CI = 0.55–0.90)) lower risk of metabolic diseases and a 5% (HR = 0.95 (95% CI = (0.70–1.31)) lower risk of malignant diseases (Table 2A and Supplementary Table 1A). In addition, Supplementary Fig. 2 shows that 50% of the LLS IPs had an age-related disease at the age of 68 years whereas this was the case at the earlier age of 65.8 years for their partners. 50% of the LLS IPs had a metabolic disease at an age of 74.8 years, while this was the case at 68.6 years for their partners, indicating a median delay of metabolic disease diagnosis for LLS IPs of 6.2 years.

IPs have a delayed onset of multimorbidity during follow-up

To study whether the delayed onset of first disease for LLS IPs extended to developing more than one disease (multimorbidity) during the 16 years of follow-up, we investigated the difference in time from inclusion to having two diseases within the same category (2 age-related, metabolic, or malignant diseases; Table 2B and Supplementary Table 1B). We observed that the yearly risk to develop 2 age-related diseases was 45% (HR = 0.55 (95% CI = 0.36–0.85) lower for the LLS IPs than for their partners, maximizing to a 49% (HR = 0.51 (95% CI = 0.26–0.97) difference for metabolic diseases. However, the yearly risk to develop 2 malignant diseases (HR = 1.39 (95% CI = 0.29-6.70)) did not significantly differ between LLS IPs and their partners. Supplementary Figure 3 shows the survival curves corresponding to Table 2B.

To study whether LLS IPs, who already had a disease, had a lower risk of getting a second disease, we investigated whether the time between first and specific second disease was longer for the LLS IPs than for their partners. Table 2C and Supplementary Table 2 show that the yearly risk to develop an age-related or a metabolic disease after already being diagnosed with an age-related disease, was 54% (HR =
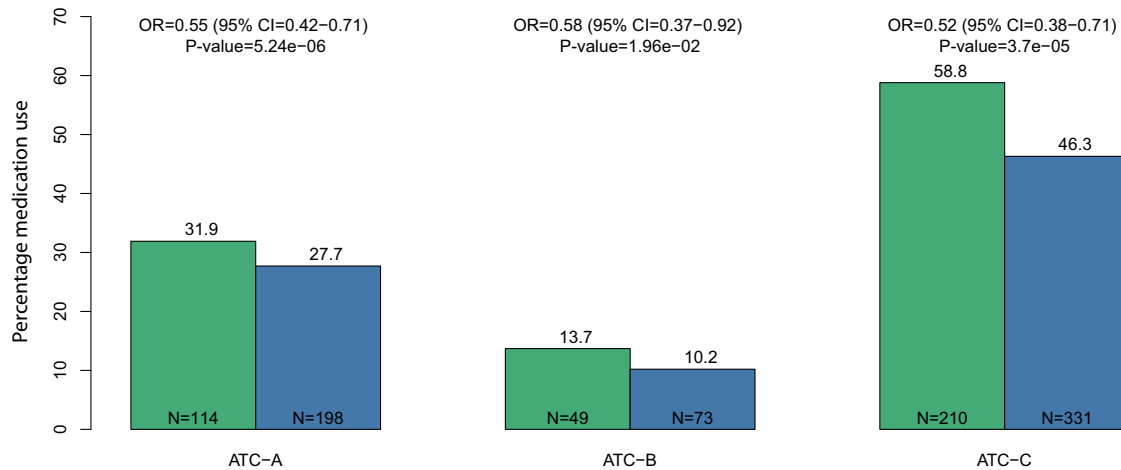
Fig. 2 | **Disease prevalence and medication use in the LLS.** This figure depicts the odds ratio's (ORs; estimated effect sizes) for medication use. Blue bars represent LLS IPs and green bars represent their partners, similar to the colors used in Fig. 1. The *y*-axis represent the percentage of LLS IPs and partners who used ATC-A (alimentary tract and metabolism; $N_{offspring} = 198$ and $N_{partners} = 114$), ATC-B (blood and blood-forming organs; $N_{offspring} = 73$ and $N_{partners} = 49$), or ATC-C (cardiovascular system; $N_{offspring} = 331$ and $N_{partners} = 210$) type medications (*x*-axis). CI is the abbreviation for confidence interval and *N* represents the numbers of the LLS IPs and partners in the specific disease groups. Statistical testing was performed using Wald tests for the Odds Ratios estimated with a mixed-model logistic regression analysis. The analyses are adjusted for age at inclusion and sex. Source data are provided as a Source Data file.

0.46 (95% CI = 0.26–0.83), and 66% (HR = 0.33 (95% CI = 0.14–0.81)) lower for the IPs, respectively. For a malignant disease following the initial diagnosis of an age-related disease, no significant difference between LLS IPs and their partners was observed (HR = 0.58 (95% CI = 0.27–1.25)). Supplementary Figure 4 shows the survival curves corresponding to Table 2C. Moreover, stratification by disease group showed that the HRs representing time from first to second disease were not affected by the group (metabolic or malignant) of first disease.

Increasing numbers of long-lived ancestors indicate a later disease onset

In our previous work, we developed the Longevity Relatives Count (LRC) score to quantify a person's number of long-lived ancestors[21,22]. The LRC score can be interpreted as a weighted proportion (ranging between 0 and 1)[22]. For example, an LRC score of 0.5 for an IP indicates 50% long-lived ancestors weighted by the genetic distance between IPs (and partners in LLS) and their ancestors. Here we investigate whether healthspan in LLS and SEDD is associated with the number of long-lived ancestors by testing whether an increasing LRC score of IPs is associated with a delay in disease onset and lower medication use in a longitudinal study design of the two independent databases; LLS and SEDD.

We conducted our analyses using two approaches. In the first approach we used the LRC score to enlarge the contrasts between the LLS IPs and their partners by defining four mutually exclusive groups: LRC_g1: IPs with an LRC ≥ 0.60, LRC_g2: IPs with an LRC [≥0.1 & <0.60], LRC_g3: partners with an LRC > 0, and LRC_g4: partners with an LRC = 0. We subsequently compared the disease incidence and medication use of LRC_g1-3 with LRC_g4, using Cox-type random effect (frailty) and linear mixed models respectively. In the second approach, we calculated the LRC score in the LLS IPs and partners combined, allowing a quantitative definition of the LRC-score instead of defining groups. Using the quantitatively defined LRC-score we investigate whether an increasing LRC score associates with a decreasing first disease incidence, using Cox-type random effect (frailty) models. Finally, we replicated the results obtained in the LLS by repeating our analysis in Swedish register data (SEDD).

First approach. When comparing the LLS IPs with an LRC score ≥0.60 (LRC_g1) with the partners who had an LRC score of 0 (LRC_g4) we observed an HR of 0.56 (95% CI = 0.34–0.92) and 0.69 (95% CI = 0.31–1.53) for time to first age-related and malignant disease, respectively (Table 3). Table 3 further shows that the healthspan benefit of the LRC_g1 group was most striking for the incidence of first metabolic disease, for which the yearly risk was 53% lower (HR = 0.47 (95% CI = 0.25–0.87)). For comparison: HR's in Table 2 (not applying LRC score) are 0.79, 0.95, and 0.71 for age-related, malignant, and metabolic diseases respectively, providing a strong indication that increasing numbers of long-lived ancestors are associated to a later disease incidence. To illustrate this comparison, Fig. 3 shows the survival curves for the LLS IPs and partners (Panel A corresponding to Table 2) and the LRC groups (Panel B corresponding to Table 3). The figure shows how the LRC score maximizes the contrast: 50% of the LRC_g4 persons developed a first metabolic disease at the age of 68 years, whereas 50% of the LRC_g1 persons developed a first metabolic disease at the age of 81 years. Hence, the LRC_g1 persons delayed the age of metabolic disease onset with a pronounced 13 years difference. The survival curves of the other disease categories are presented in Supplementary Fig. 5. Further benefit for LRC_g1 over LRC_g4 concerns the development of multimorbidity and medication use, since an HR of 0.14 (95% CI = 0.03–0.70) was observed for the time to develop 2 age-related diseases and an OR of 0.26 (95% CI = 0.12–0.57) for medication use.

Second approach. We calculated the LRC score for the LLS IPs and their partners combined to avoid any grouping. Table 4A shows that with every 0.1 (10%) increase in LRC score, LLS F3 participants had a 5% (HR = 0.95 (95% CI = 0.91–0.99)) lower yearly risk to develop a first age-related disease. To illustrate the magnitude, this effect increases to 50% when all ancestors were long-lived (LRC score = 1). We further observed a 7% (HR = 0.93 (95% CI = 0.88–0.98)) lower yearly first metabolic and 3% (HR = 0.97 (0.91–1.04)) lower malignant disease risk, though the latter effect was not statistically significant.

We replicated the results obtained in the LLS by repeating our analyses in Swedish register data (SEDD). Table 4B shows that with every 10% increase in LRC score, the SEDD IPs have a 6% (HR = 0.94 (95% CI = 0.89–0.98)) lower yearly risk to develop a first age-related disease, 9% (HR = 0.91 (0.87–0.96)) lower risk for metabolic and 5% (HR = 0.95 (0.90–0.99)) for malignant disease. Moreover, the yearly risk of dying decreases 8% (HR = 0.92 (0.87–0.97)) with every 10% increase in LRC score.

Table 2 | LLS disease incidence

	N (prop)	Events (prop)	HR (95% CI)	P-value
A: Time from inclusion to first disease				
Age-related diseases				
LLS IPs	917 (0.70)	362 (0.39)	0.79 (0.65–0.97)	2.32 × 10⁻²
Partners (ref)	395 (0.30)	171 (0.43)	–	–
Metabolic diseases				
LLS IPs	917 (0.70)	261 (0.23)	0.71 (0.55–0.90)	5.18 × 10⁻³
Partners (ref)	395 (0.30)	135 (0.34)	–	–
Malignancies				
LLS IPs	917 (0.70)	130 (0.14)	0.95 (0.70-1.31)	7.66 × 10⁻¹
Partners (ref)	395 (0.30)	56 (0.14)	–	–
B: Time from inclusion to 2 diseases				
Age-related diseases				
LLS IPs	611 (0.70)	73 (0.12)	0.55 (0.36–0.85)	6.83 × 10⁻³
Partners (ref)	268 (0.30)	47 (0.18)	–	–
Metabolic diseases				
LLS IPs	611 (0.70)	45 (0.07)	0.51 (0.26–0.97)	3.96 × 10⁻²
Partners (ref)	268 (0.30)	29 (0.11)	–	–
Malignancies				
LLS IPs	611 (0.70)	7 (0.01)	1.39 (0.29-6.70)	6.82 × 10⁻¹
Partners (ref)	268 (0.30)	2 (0.01)	–	–
C: Time from first disease to second disease				
Age-related → Aage-related diseases				
LLS IPs	500 (0.68)	79 (0.16)	0.46 (0.26–0.83)	9.82 × 10⁻³
Partners (ref)	237 (0.32)	55 (0.23)	–	–
Age-related → Metabolic diseases				
LLS IPs	500 (0.68)	62 (0.12)	0.33 (0.14–0.81)	1.46 × 10⁻²
Partners (ref)	237 (0.32)	44 (0.19)	–	–
Age-related → Malignant diseases				
LLS IPs	500 (0.68)	17 (0.03)	0.58 (0.27-1.25)	1.66 × 10⁻¹
Partners (ref)	237 (0.32)	11 (0.05)	–	–

Table shows the time from inclusion to first disease in panel A, the time from inclusion to having 2 diseases (panel B), and the time from first to second disease (panel C). N is the group size used for the analyses and prop. is the proportion from the total. Events are the events of the specific diseases, for example, age-related diseases, and prop. indicates the proportion from the size of a specific group (LLS IPs or partners). For example, in panel A, there are 917 offspring and 395 partners. The total is 1312. 917 is 70% out of the total and 395 is 30% of the total. 362 events are then 39% out of the 917 offspring and 171 events are 50% out of the 395 partners. HR is the abbreviation for Hazard Ratio and represents the estimated effect size. Statistical testing was performed using Wald tests for the conditional log-hazard ratio estimated with a Cox-type frailty regression model. The analyses are adjusted for sex, different ages of study entry (left truncation), and right censoring. Survival curve details can be found in Supplementary Figs 2–4. To study time from inclusion to first disease, only persons without any disease at inclusion were investigated. Note that the numbers in panel B are lower than those in panel A. This is because the censoring group reflects those for whom we have not observed any disease at the end of follow-up. As a result, persons with only one disease are excluded from the analyses. Moreover, the analyses in panel C are confined to persons who experienced a first disease. The numbers are larger than the events reported in panel A because we allowed for 1 disease prior to the start of follow-up (see "Methods" section for more details).

IPs already had a healthy metabolomic profile at inclusion

Our results point strongly towards protection from metabolic diseases for persons with an increasing number of long-lived ancestors as established with a high LRC score. We therefore investigated whether those with a high LRC score at the moment of inclusion in the LLS, indeed have a healthy circulating metabolomic profile that marks protection from disease at midlife. To estimate health as a quantitative parameter, we used a recently developed NMR-metabolomics-based predictor of 5–10 years all-cause mortality across all ages from midlife onwards (from here MetaboHealth score)[25]. Hence, we explored whether the MetaboHealth score associates with differences between LRC groups as defined in the first approach of the analysis above (LRC_g1-3

compared to LRC_g4) and conducted a mixed model linear regression analysis.

We observed that the IPs with an LRC score ≥0.60 (LRC_g1 IPs) had a 0.098 (95%CI: [−0.184] – [−0.012]) lower MetaboHealth score than the partners who had an LRC score of 0 (LRC_g4 IPs; Fig. 4 and Supplementary Table 3). The LRC_g2 and LRC_g3 IPs had a 0.032 (95%CI: [−0.077] – [0.012]) and a 0.016 (95%CI: [−0.091] – [0.058]) lower score than the LRC_g4 IPs respectively. Though the effects are relatively small (Fig. 4), indeed we observed that the LLS IPs with ≥60% long-lived ancestors, who show delayed onset of disease, also have a healthier circulating metabolomic profile in mid-life than the partners with an LRC score of 0.

Discussion

Human longevity is heritable and clusters in specific families. Members of these families live longer and seem to age healthier than the general population. Studying these long-lived families is important to improve our understanding of the molecular and environmental mechanisms that protect from (multi)morbidity and promote a healthy survival up to high ages. In this study we investigated the development of diseases from mid-life onwards in big multigenerational and prospective data, covering up to 26 years of follow-up, in family based (LLS, Netherlands) and population based (SEDD, Sweden) data. We showed that members of long-lived families have a delayed onset of disease, multimorbidity, and medication use as compared to their partners, thereby extending their healthspan with up to a decade. These members also postponed multimorbidity since those who were already diagnosed with an age-related disease (Supplementary Table 4) had a 54% lower risk of having a second age-related disease compared to their partners. When defining familial longevity quantitatively using the LRC score, we demonstrated that an increasing number of long-lived ancestors associates with an increasing delay in disease onset and lower medication use. Finally, we demonstrated that at the moment of LLS study inclusion, those with an LRC score ≥0.60 (LRC_g1)) had a better MetaboHealth score than their partners with an LRC score of 0 (LRC_g4), indicating better immune and metabolic health, and lower 5-years mortality risk. We conclude that an increasing number of long-lived ancestors, as measured with the LRC score, is a quantitative indicator of familial longevity, capturing decreased mortality risk, protection against multimorbidity, and improved health in selected families as well as the population. The LRC score can thus potentially be used in genetic studies to elucidate multimorbidity-limiting mechanisms that promote healthspan already in mid-life.

Our analyses, using ancestral mortality data, in the selected Dutch longevity families and the Swedish register data led to remarkably similar conclusions. An increasing number of long-lived ancestors, as measured with the LRC score, not only associates with a lower mortality at any moment in life[21,22], it also associates, in a similar way, with a lower disease incidence during mid and later life (60–75 years): With every 10% increase in LRC score the yearly risk to develop an age-related disease decreased with 5% in the LLS, and 6% in the SEDD, maximizing to 50% and 60% respectively when all ancestors were long-lived.

We did observe stronger effects in the SEDD data than in the LLS data, with consistently lower hazard ratio's (HRs) for age-related and metabolic disease incidence. This may be explained firstly because LLS IPs are compared with their partners as controls, either as separate or combined groups in the LRC analyses. IPs and partners share the same adult household and thus, the LLS design controls for shared resources and behavior (such as socio-economic status, social network, and lifestyle). In the SEDD data we did not compare IPs with their partners. The effect size difference between LLS and SEDD may therefore represent the influence of shared resources and behavior[26]. Secondly, in the LLS disease diagnoses were obtained from the general practitioners (GPs) whereas in the SEDD, disease diagnoses were obtained

Table 3 | LLS disease incidence and medication use in LLS LRC groups

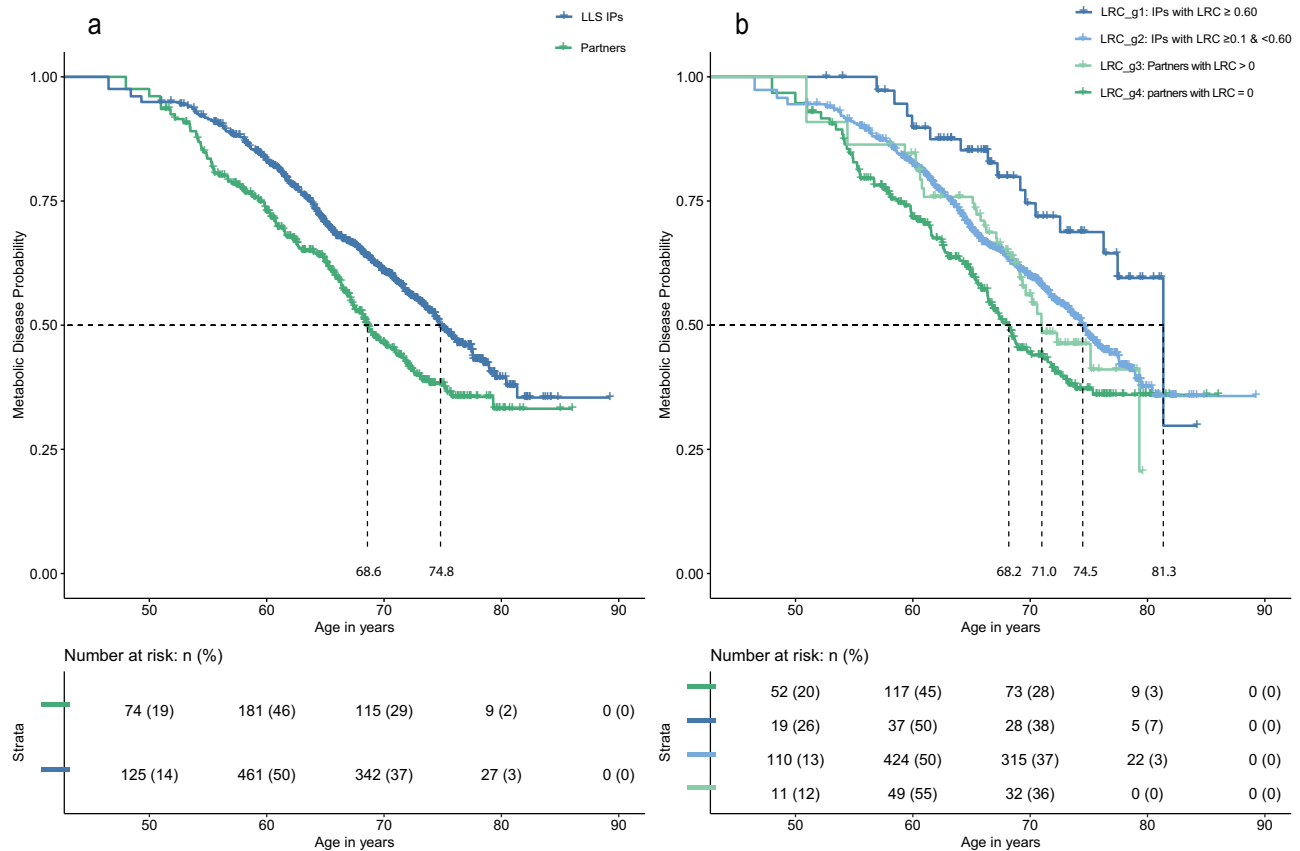| Panel A: disease incidence				
A: Time from inclusion to first disease				
	N (prop)	Events (prop)	HR (95% CI)	*P*-value
Age-related diseases				
LRC_g4: partners with LRC = 0 (ref)	262 (0.21)	113 (0.43)	–	–
LRC_g3: partners with LRC > 0	89 (0.07)	38 (0.43)	0.85 (0.58–1.26)	4.29×10^{-1}
LRC_g2: IPs with LRC ≥ 0.1 & <0.60	843 (0.66)	339 (0.40)	0.80 (0.63–1.01)	5.98×10^{-2}
LRC_g1: IPs with LRC ≥ 0.60	74 (0.06)	23 (0.31)	0.56 (0.34–0.92)	2.17×10^{-2}
Metabolic diseases				
LRC_g4: partners with LRC = 0 (ref)	262 (0.21)	88 (0.34)	–	–
LRC_g3: Partners with LRC > 0	89 (0.07)	30 (0.34)	0.82 (0.51–1.30)	3.89×10^{-1}
LRC_g2: IPs with LRC ≥ 0.1 & <0.60	843 (0.66)	246 (0.30)	0.72 (0.54–0.95)	2.17×10^{-2}
LRC_g1: IPs with LRC ≥ 0.60	74 (0.06)	15 (0.20)	0.47 (0.25–0.87)	1.72×10^{-2}
Malignancies				
LRC_g4: partners with LRC = 0 (ref)	262 (0.21)	37 (0.14)	–	–
LRC_g3: Partners with LRC > 0	89 (0.07)	14 (0.16)	1.04 (0.56-1.94)	8.93×10^{-1}
LRC_g2: IPs with LRC ≥ 0.1 & <0.60	843 (0.66)	122 (0.13)	0.97 (0.64-1.46)	8.89×10^{-1}
LRC_g1: IPs with LRC ≥ 0.60	74 (0.06)	8 (0.11)	0.69 (0.31-1.53)	3.62×10^{-1}
B: Time from inclusion to 2 diseases				
Age-related diseases				
LRC_g4: partners with LRC = 0 (ref)	177 (0.21)	31 (0.18)	–	–
LRC_g3: Partners with LRC > 0	60 (0.07)	9 (0.15)	0.66 (0.28-1.54)	3.33×10^{-1}
LRC_g2: IPs with LRC ≥ 0.1 & <0.60	565 (0.67)	71 (0.13)	0.56 (0.34–0.93)	2.50×10^{-2}
LRC_g1: IPs with LRC ≥ 0.60	46 (0.05)	2 (0.04)	0.14 (0.03–0.70)	1.65×10^{-2}
Panel B: medication use				
	N (prop)	Events (prop)	OR (95% CI)	*P*-value
ATC-C (metabolic medication)				
LRC_g4: partners with LRC = 0 (ref)	428 (0.21)	144 (0.35)	–	–
LRC_g3: Partners with LRC > 0	149 (0.07)	43 (0.29)	0.70 (0.39-1.25)	2.26×10^{-1}
LRC_g2: IPs with LRC ≥ 0.1 & <0.60	1363 (0.66)	306 (0.23)	0.43 (0.30–0.63)	1.02×10^{-5}
LRC_g1: IPs with LRC ≥ 0.60	124 (0.06)	24 (0.19)	0.26 (0.12–0.57)	7.16×10^{-4}

In this table the analyses with the largest sample size are repeated after defining new groups based on the ancestral mortality information of the LLS IPs and partners. LLS IPs were separated into two groups. Group 1 are those who have 60% of their ancestor belonging to the top 10% of their birth cohort (LRC_g1) and group 2 are the other, original LLS, offspring (LRC_g2). Partners were also separated into two groups. Group 3 are those who have at least one parent that belongs to the top 10% survivors of his/her birth cohort (LRC_g3) and group 4 are those without any parents who belong to the top 10% survivors of their birth cohort (LRC_g4). As some ancestors of LLS IPs were still alive, they were excluded from the LRC score (see methods for more details). 236 ancestors of partners were still alive. To not lose those who were still alive for our analyses, we used single imputation to predict the remaining life expectancy after having reached a specific age (last age of observation) based on the Dutch population life expectancy. Lifetables were provided by Statistics Netherlands from 1850 up to 2021 (see methods for more details). HR is the abbreviation for Hazard Ratio and represents the estimated effect size. Statistical testing was performed using Wald tests for the conditional log-hazard ratio estimated with a Cox-type frailty regression model. The analyses are adjusted for sex, different ages of study entry (left truncation) and right censoring.

from hospital records available in the Swedish national register data. It may be that stronger effects were observed in the SEDD because hospitalization is on average an indication of more extreme health events than receiving a GP diagnosis. Nevertheless, for many of the GP diagnosed diseases, such as a myocardial infarction, hospitalization is also required.

We did not observe statistically significant results for malignancies in the LLS data whereas we did observe significant effects in the SEDD data. However, within the SEDD data the effects for malignant disease incidence were considerably smaller (HR closer to reference group) than for metabolic and age-related diseases. A first explanation for this observation, which also applies to the generally smaller effects in the LLS compared to SEDD, relates to differences in study population and follow-up time. LLS IPs and partners were followed-up for a maximum of 16 years from the average age of 59 years whereas for the SEDD IPs this was a maximum of 26 years from the average age of 52 years. As a result, the protecting effect of familial longevity may be a bit more mitigated in the LLS[27]. Moreover, we may have missed early onset (~50 years) malignancies in the LLS whereas those are included in the SEDD. This could also explain why a

considerably lower proportion of malignant diseases was available in LLS than in the SEDD whereas this was not the case for metabolic diseases. Secondly, inherited genetic factors have a limited effect on many types of malignancies, with heritability estimates ranging between 20% and 30%[28]. However, the chronic diseases, as measured in our study, are much more heritable[29–38], with over 70% heritability for type 2 diabetes[39,40]. As the LRC score captures additive genetic effects[21,22], the lower heritability of malignancies could explain why the small number of malignant disease observations in the LLS did not provide enough power to detect effects and why the effect sizes are lower in the SEDD compared to metabolic diseases. Previous research focusing on malignant diseases in long-lived individuals and their offspring obtained heterogenous results[4,5,7,8,10,11,13–15,17,41–45] which may also be due to study population and selection differences.

Past research primarily focused on studying disease prevalence of long-lived individuals, such as centenarians, and their children[4,9–14,17–19] in cross-sectional designs. Our data covers up to 26 years of follow-up and provides a unique combination between ancestral mortality information and deep phenotyping of chronic age-related diseases, medication use, as well as metabolomics. This allowed us to closely link

Fig. 3 | LLS metabolic disease incidence with and without LRC-defined groups. This figure depicts survival curves reflecting metabolic disease incidence within the Leiden Longevity Study (LLS). The *x*-axis show age in years and the *y*-axis show metabolic disease incidence. Dotted lines represent the age at which 50% of the members of a specific group had their first metabolic disease. **a** depicts two groups; the blue line represents LLS Index Persons (IPs; $N = 917$) and the green line represents the partners ($N = 395$). The mean difference between the lines represents the Hazard Ratio (HR; estimated effect size) shown in Table 2. **b** depicts four groups;

LRC_g1: IPs with an LRC ≥ 0.60 (dark blue; $N = 74$), LRC_g2: IPs with an LRC [≥0.1 & <0.60] (light blue; $N = 843$), LRC_g3: partners with an LRC > 0 (light green; $N = 89$), and LRC_g4: partners with an LRC = 0 (dark green; $N = 262$). The mean difference between the LRC_g1-3 and LRC_g4 line represents the HRs shown in Table 3. Vertical lines within the colored lines represent right censored events. The bottom column of panel a and b shows how many persons were still at risk of having a metabolic disease at different ages. Survival curves are adjusted for left truncation and right censoring. Source data are provided as a Source Data file.

Table 4 | Quantitative LRC analyses of time to first disease in LLS en SEDD

	N	Events (prop)	HR (95% CI)	*P*-value
A: LLS data				
Age-related diseases	1312	533 (0.41)	0.95 (0.91–0.99)	3.06×10^{-2}
Metabolic diseases	1312	396 (0.30)	0.93 (0.88–0.98)	1.37×10^{-2}
Malignancies	1312	186 (0.14)	0.97 (0.91-1.04)	4.98×10^{-1}
B: SEDD data				
Age-related diseases	2497	1,190 (0.48)	0.94 (0.89–0.98)	1.22×10^{-3}
Metabolic diseases	2497	706 (0.28)	0.91 (0.87–0.96)	4.84×10^{-3}
Malignancies	2497	671 (0.27)	0.95 (0.90–0.99)	4.43×10^{-2}
Death	2497	694 (0.28)	0.92 (0.87–0.97)	1.72×10^{-3}

Table shows the time from inclusion to first disease. *N* is the group size. Events are the events of the specific diseases, for example, age-related diseases, and prop. indicates the proportion from the size of a specific disease group. for example, 533 (41%) out of the 1312 persons had an age-related disease. HR represents the estimated effect size. Statistical testing was performed using Wald tests for the conditional log-hazard ratio estimated with a Cox-type frailty regression model. The analyses are adjusted for sex, different ages of study entry (left truncation) and right censoring. Only persons without any disease at inclusion are studied.
sd standard deviation, *HR* hazard ratio.

familial longevity to medication use and the incidence of multiple diseases. Detailed information about disease incidence was provided by the treating physicians (General Practitioners, GPs) of the LLS IPs and their partners. In the SEDD, we used hospitalization records from the Swedish national registers (see "Methods" section for more details). The combination between these two types of data ensured

robustness against healthy participant bias. Together, the LLS and Swedish data cover a very large and robustly measured (GP records in the Netherlands and hospitalization records in Sweden) group of persons to investigate disease incidence from a familial perspective, with up to 26 years of disease follow-up. Apart from this, the LLS was initiated with the inclusion of (1) LLS IPs who had at least one long-lived
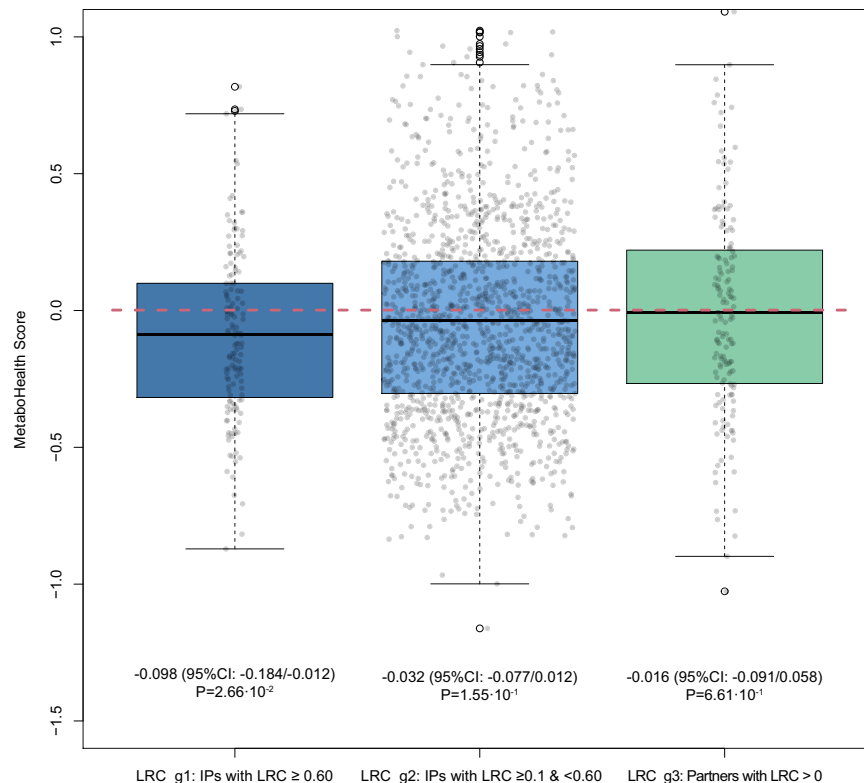
Fig. 4 | MetaboHealth score differences for LRC groups at LLS study inclusion. The *x*-axis depicts three groups; LRC_g1: IPs with an LRC ≥ 0.60 (dark blue; *N* = 121), LRC_g2: IPs with an LRC [≥0.1 and <0.60] (light blue; *N* = 1297), and LRC_g3: partners with an LRC > 0 (light green; *N* = 135). The dotted red line depicts the mean MetaboHealth score of the LRC_g4 group: partners with an LRC = 0 (dark green; *N* = 397). The *y*-axis depicts the MetaboHealth score. Higher MetaboHealth score values represent a less healthy metabolomic profile as measured by the MetaboHealth score. The score represents 5/10 year mortality risk (see "Methods" for more details). The middle line in the boxes represents the mean, the edges of the boxes represent the first quartile and the extreme whiskers represent the third quartile. The bottom of the figure depicts the beta coefficients (estimated effect sizes) for the comparison between LRC_g1-3 with LRC_g4 and correspond to Supplementary Table 3. Statistical testing was performed using T tests for the Beta coefficients estimated with a mixed-model linear regression analysis. The analyses are adjusted for sex, age at inclusion, and medication use. Source data are provided as a Source Data file.

parent and aunt or uncle and (2) the partners of the LLS IPs. LLS IPs are likely to share social (e.g. social network, socio-economic status) and behavioral (e.g. lifestyle, drinking, sporting) traits important for healthy aging and longevity, for example, because they share the same household or due to assortative mating[26]. The LLS study design thus corrects for such similarities between LLS IPs and partners, potentially resulting in an underestimation of differences between LLS IPs and partners as compared to general population controls[9]. However, replication in the SEDD, which does not contain any initial inclusion criteria, guarantees results which are not influenced by partner similarities.

Genetic longevity studies so far mainly focused on survival to exceptional ages. Using the LRC score, disease-free survival, possibly in combination with the MetaboHeath score, may be explored as a broader phenotype to increase the power of longevity genetic studies. In addition, the association between LRC score and delayed disease incidence may be explained by the presence of variants protecting from development of disease and/or the absence of disease loci in the long-lived families. Though, previous research showed no evidence that long-lived persons were characterized by the absence of disease loci[46], GWAS studies identifying disease susceptibility variants for example, for hypertension[47], Alzheimer's disease[48] has progressed significantly. Hence, it is interesting to re-investigate if the absence of disease susceptibility loci associates to the LRC score. As mentioned in the previous paragraphs, the larger effect sizes in the SEDD likely illustrate the importance of shared resources and behavior in long-lived families. Further evidence for the clustering of socio-behavioral traits was provided in a recent study which showed that members of

long-lived families were less frequently hospitalized with smoking-related cancers as a first disease[8]. As socio-behavioral traits are influenced by complex combinations between genes and environment, further investigation may aid genetic research while providing an interesting basis to investigate the social complexity underlying familial longevity.

Our results provide strong evidence that an increasing number of long-lived ancestors associates with up to a decade of healthspan extension and a healthy metabolomic profile in mid-life. Our results have two important implications. First, future genetic research aimed at identifying protective longevity mechanisms beneficially influencing the risk of multimorbidity could focus on a broader definition of longevity entailing survival to exceptional ages as well as disease-free survival and possibly the MetaboHealth score metabolites. Second, our results highlight the importance of integrating multiple generations of ancestral mortality data to existing and novel studies. In the past it was difficult to obtain such ancestral information but currently it is much more feasible to do so, as population scale family tree data is becoming increasingly available[21,49,50]. Moreover, in an increasing number of countries ancestral data can be retrieved from the national statistics bureaus, such as Statistics Sweden or Statistics Netherlands. Finally, next to genetic drivers of longevity and disease incidence, it is important to investigate if and how potential socio-behavioral resources[51], such as socio-economic status and stress, associate to both longevity and disease incidence. If these novel insights are consistently applied across studies, the comparative nature of longevity studies may improve and facilitate the discovery of novel genetic variants and mechanisms promoting healthy ageing.

Methods

Leiden Longevity Study

The Leiden Longevity Study (LLS) was initiated in 2002 to study the mechanisms that lead to exceptional survival. Inclusion took place between 2002 and 2006 and initially started with the recruitment of living sibling pairs. Within a sibling pair, males were invited to participate if they were 89 years or older and females if they were 91 years or older. Inclusion was subsequently extended to the children of the sibling pairs and their partners. In the current study, we focus on the children of the sibling pairs and their partners, referring to them as LLS IPs and partners. From their perspective, IPs were included if they had at least one long-lived parent and aunt or uncle (females ≥ 91 years and males ≥ 89 years). In total, 1674 Index Persons (IPs, F3), 745 partners (F3), 1295 parents (F2), 2370 aunts and uncles (F2), 760 grandparents (F1), and 1237 parents of the partners (F2) were included in this study. The LLS currently consists of 651 three-generational families, defined by IP siblings who have the same parents (Fig. 1).

Mortality information was verified by birth or marriage certificates and passports whenever possible. Additionally, verification took place via personal cards which were obtained from the Dutch Central Bureau of Genealogy. In January 2021 all mortality information was updated through the Personal Records Database (PRD) which is managed by Dutch governmental service for identity information. https://www.government.nl/topics/personal-data/personal-records-database-brp. The combination of officially documented information provides very reliable and complete ancestral as well as current mortality information.

Disease data has been retrieved from the General Practitioners (GPs) of the LLS IPs and partners and covers the period from birth until 2018. GPs extracted the presence of chronic age-related diseases as specified in Supplementary Table 4 and the year the disease occurred from their electronic health records. The GP records are kept up to date when a person switches from one GP practice to another. Diseases were clustered into 3 groups based on the International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes, (1) metabolic diseases, (2) malignant diseases, and (3) age-related diseases, which are the combination of metabolic and malignant diseases. Furthermore, cross-sectional information on medication use from pharmacies was obtained for the period 2006-2008, indicating whether a specific medicine was used or not. Medication was grouped according to the international Anatomical Therapeutic Chemical Classification System (ATC) standard. We focused on ATC-A (alimentary tract and metabolism), ATC-B (blood and blood forming organs), and ATC-C (cardiovascular system) type medications because they match the disease groups we investigate.

Ethylenediamine tetraacetic acid (EDTA) plasma samples were obtained for all LLS IPs and partners at inclusion. From these samples, metabolomics biomarker data was quantified using high-throughput nuclear magnetic resonance (NMR) spectroscopy provided by the Nightingale Health platform. Experimentation and application details of the Nightingale NMR platform has previously been described[52,53]. Moreover, the metabolic biomarkers measured using the nightingale platform were used in a variety of publications (overview can be found here: https://nightingalehealth.com/publications).

Scanian Economic-Demographic Database

The Scanian Economic-Demographic Database (SEDD) is a longitudinal database covering five rural Scanian parishes and the city of Landskrona. It spans the period 1812-1967, with full coverage of the villages from 1812 and for Landskrona from 1904. The SEDD database was constructed using register-type data from catechetical examination registers and was updated with information on births, marriages, and deaths from church books. Unique person numbers were introduced in Sweden by 1947. Through these person numbers individuals can be followed in the national Swedish registration, introduced in 1968. Persons who out-migrated from the research region before the introduction of the person number were linked to the 1950 Census and the Swedish Death Index. The obtained person numbers were subsequently used to track individuals in the Swedish national register for the period 1968-2015. The link to the Swedish Death Index yielded ancestral death dates anywhere in Sweden even for individuals who out-migrated from the research region before the person number or nationwide register data were introduced. At present (2023), the SEDD database contains 920,159 unique individuals.

Index person (IP) identification for this study happened in subsequent steps (Supplementary Table 5). First, from the entire SEDD data we identified all persons (from here: IPs) who were part of the national register data in the years 1990–1995 and between ages 45–60, and followed them in the national registers for the period 1990–2015. Second, IPs were selected to have known grandparents on at least one side of the family (maternal or paternal), and whose parents were from an extinct birth cohort (born before 1915) to ensure complete information about their date of death. Third, we included lifespan information of their parents, aunts and uncles, and their grandparents. Fourth, IPs who were found in the hospital records in the year preceding their eligibility for the study (1989–1994) were excluded to minimize the number of IPs with existing conditions receiving hospital treatments. Lastly, partners of IPs were excluded to ensure mutually exclusive ancestral information. In total, 1493 Index Persons (IPs, F3), 2969 parents (F2), 5830 aunts and uncles (F2), and 3028 grandparents (F1) were included in this study. The SEDD consists of 1495 three-generational families, defined by IP siblings who have the same parents (Fig. 1).

The Swedish hospital registers reached nationwide coverage in 1987 and records are considered complete from 1989. The main diagnosis for each hospitalization has been recorded in ICD-9 coding from 1987 to 1997 and ICD-10 coding 1997 to 2015. We recoded ICD-9 diagnoses to ICD-10 using the official crosswalk provided by Statistics Sweden. Diseases are specified identical to the LLS (Supplementary Table 4) to ensure comparability between the databases. It is relevant for our analyses to mention that only 214 IPs (8.6%) die without ever receiving a hospital diagnosis as a higher percentage would have warranted a competing risk analysis (see statistics section for more details).

Lifetables

In the Netherlands and Sweden, population-based cohort lifetables are available from 1850 and 1800 respectively, until 2021[54,55]. These lifetables contain, for each birth year and sex, an estimate of the hazard of dying between ages x and x + n (hx) based on yearly intervals ($n = 1$) up to 99 years of age. Conditional cumulative hazards (Hx) and survival probabilities (Sx) can be derived using these hazards. In turn, we can determine to which sex and birth year based survival percentile each person of our study belonged to. For example: a person was born in 1876, was a female, and died at age 92. According to the lifetable information this person belonged to the top three percent survivors of her birth cohort, meaning that only three percent of the women born in 1876 reached a higher age. We used the lifetables to calculate the birth cohort and sex-specific survival percentiles for all persons in the LLS and SEDD. This approach takes sex-specific differences in longevity into account and prevents against the effects of secular mortality trends over the last centuries and enables comparisons across study populations[56,57]. In SEDD, we focused only on extinct birth cohorts and death ancestors. However, In the LLS some ancestors (only aunts/uncles) were still alive (right censoring). To deal with non-extinct birth cohorts, we used the prognostic lifetables provided by Statistics Netherlands[54,55] and to deal with right censoring we used single imputation where we estimated an age of death based on the remaining life expectancy at the age of censoring.

Scores

LRC score. The Longevity Relatives Count (LRC) score was used in LLS and SEDD to map the IPs' family history of longevity. The LRC score

indicates the proportion of ancestors that became long-lived, weighted by the genetic distance between IPs (and partners in LLS) and their ancestors. For example, an LRC of 0.5 indicates 50% long-lived ancestors. For this study, two generations of ancestors were available to calculate the LRC score for IPs and one generation for the partners of the LLS IPs (Fig. 1 and Supplementary Fig. 1). In the LLS, the LRC score was calculated using the mortality information as updated in 2021. In the SEDD, IPs were identified in such a way that all ancestors were deceased at the start of follow-up. The LRC score has been described in detail by van den Berg et al., 2020 in Aging Cell[22] and can be calculated as follows:

$$LRC_i = \frac{\text{weighted number of top10\%ancestors}}{\text{weighted total number of ancestors}} = \frac{\sum_{k=1}^{N_i} w_k \cdot I(P_k \geq 0.9)_i}{\sum_{k=1}^{N_i} w_k}$$

Where i refers to the IPs for whom the score is built. k is an index referring to each ancestral blood relative (from here: ancestors) of person i who are used to construct the score. N_i refers to the total number of ancestors of person i, Fig. 1), P_k is the sex and birth year-specific survival percentile, based on lifetables, of ancestor k, and $I(P_k \geq 0.9)$ indicates if ancestor k belongs to the top 10% survivors. $\sum_{k=1}^{N_i} w_k$ is the weighted total number of ancestors of IP i. The relationship coefficients are used as weights w_k. Supplementary Figure 6A-B depict the LRC score distribution in the LLS and SEDD.

MetaboHealth score. In de LLS study, the MetaboHealth score was used as an indicator for the (metabolomic) health of the offspring and partners at study inclusion. This previously published score was generated based on NMR metabolomics data in ~40.000 European study participants and is calculated as the weighted sum of 14 independent metabolites covering 5–10 years mortality risk and metabolite markers of lipid metabolism, fatty acid metabolism, glycolysis, fluid balance, and inflammation[25,58]. Supplementary Figure 6C depicts the MetaboHealth score distribution in the LLS.

Statistical analyses
Statistical analyses were conducted using R version 4.0.2[59]. We reported 95% confidence intervals (CIs) and considered two-sided p-values statistically significant at the 5% level (α = 0.05). A list of used R-packages and version numbers is available on gitlab (see code availability statement). Linear, logistic, and Cox-type random effects models were used to adjust for within-family relations, assuming family-specific random effects and defining families as F3 IPs who share the same parents.

Logistic and linear mixed models. Analysis 1: To compare medication prevalence between (LRC-based) LLS IPs and partners we fitted a logistic mixed-model:

$$\text{logit}\left(\boldsymbol{\pi}_{ij}\right) = \beta Z_{ij} + \boldsymbol{\gamma} \boldsymbol{X}_{ij} + u_i, \tag{1}$$

where Y_{ij} is the binary response (medication yes/no) for person j in family i and $\pi_{ij} = P\left(Y_{ij} = 1 | Z_{ij}, \boldsymbol{X}_{ij}, u_i\right)$ is the probability of medication conditional on the family-specific random effect u_i and individual-specific covariate information $(Z_{ij}, \boldsymbol{X}_{ij})$, β is the main regression coefficient of interest, representing the fixed effect of the variable Z which indicates a family history of longevity. We considered three types of medication resulting in three different definitions of the outcome Y: (1) ATC-A, (2) ATC-B, and (3) ATC-C type medication and two different definitions of Z: (1) offspring/partners or (2) LRC-based offspring/partners. $\boldsymbol{\gamma}$ is a vector of regression coefficients for the fixed effects of possible confounders ($\boldsymbol{X}$) which are: (1) age at study inclusion and (2) sex in all models. The unobserved family-specific random effects u_i were assumed to follow a normal distribution.

Analysis 2: To compare the MetaboHealth score between the LRC groups (LRC_g1-3 with LRC_g4) in the LLS we fitted a linear mixed-model:

$$Y_{ij} = \beta Z_{ij} + \boldsymbol{\gamma} \boldsymbol{X}_{ij} + u_i, \tag{2}$$

where Y_{ij} is the MetaboHealth score for person j in family i. β is the main regression coefficient of interest, representing the fixed effect of the variable Z which is the LRC-based offspring/partners. $\boldsymbol{\gamma}$ is a vector of regression coefficients for the fixed effects of possible confounders ($\boldsymbol{X}$) which are: (1) age at study inclusion, (2) sex, and (3) medication use. The unobserved family-specific random effects u_i were assumed to follow a normal distribution.

Survival analysis (Cox-type random effects regression model). Analysis 3: To compare age at disease onset between (LRC-based) offspring and partners in the LLS, we fitted a set of Cox-type random effects modes:

$$\lambda\left(t_{ij}\right) = u_i \lambda_0\left(t_{ij}\right) \exp(\beta Z_{ij} + \boldsymbol{\gamma} \boldsymbol{X}_{ij}) \tag{3}$$

where t_{ij} is the right-censored outcome of IP j in family i, $\lambda_0\left(t_{ij}\right)$ refers to the baseline hazard, which is left unspecified, and β is the main regression coefficient of interest, representing the fixed effect of the variable Z which indicates a family history of longevity. We have considered two alternative definitions of the outcome variable t: (1) age at first disease onset, (2) age at the second disease onset and three different definitions of Z: (1) offspring/partners, (2) LRC-based offspring/partners, and (3) LRC score as a continuous variable. The models are adjusted for left truncation given by the age at entry in the study and adjusted by sex (confounder denoted by X in expression 3). $u_i > 0$ refers to an unobserved random effect (frailty) shared by the members of the same family i and was assumed to follow a gamma distribution.

Analysis 4: To compare time between first and second disease between (LRC-based) offspring and partners in the LLS, we fitted a Cox-type random effects model (see expression 3). Here t_{ij} is age at second disease onset where age at the first disease onset is considered as the left-truncation time in this analysis. $\lambda_0\left(t_{ij}\right)$ refers to the baseline hazard, which is left unspecified in a Cox-type model. β is a vector of regression coefficients for the main effects of interest (Z) which are: (1) offspring/partners, (2) LRC-based offspring/partners, and (3) LRC score. $\boldsymbol{\gamma}$ is a vector of regression coefficients for the effects of possible confounders ($\boldsymbol{X}$) which are: (1) sex and (2) age, as captured by adjusting for left truncation. $u_i > 0$ refers to an unobserved random effect (frailty) shared by the members of the same family i and was assumed to follow a gamma distribution.

Analysis 5: Disease incidence analyses in the LLS are repeated in the SEDD data. The main goals of replication were: (1) reduction of chance findings; two independent studies showing comparable results imply lower chance of spurious associations, (2) generalizability to unselected samples; the LLS may be exposed to (healthy) participant bias and other inclusion criteria related selections while in the Swedish data does not apply any inclusion criteria and comes from the national registers which cover the entire country, (3) robustness with respect to missing information; while the LLS presents limited ancestral mortality data for partners, the Swedish data is based on (National) registers where missing information is negligible. In the SEDD analysis, we consider Cox-type random effect models following expression (3) and considering different definitions for the outcome t: (1) age at first disease onset, and (2) age at death. The models are adjusted for left truncation given by the age at entry in the study. $\lambda_0(t_{ij})$ refers to the baseline hazard, which is left unspecified and β is the regression coefficient of interest representing the fixed effect of variable Z which is the LRC score (considered as a continuous variable). As in the LLS-based Analysis 3, the models are adjusted for left truncation given by

the age at entry in the study and adjusted by sex (confounder denoted by X in expression (3)). $u_i > 0$ refers to an unobserved random effect (frailty) shared by the members of the same family i and was assumed to follow a gamma distribution.

In all Cox-type random effects regression analyses we accounted for the fact that follow-up did not start at the same age for most IPs (left truncation) by using an IP specific age at start of follow-up. Moreover, we accounted for the fact IPs may still be at risk of disease onset after the end of follow-up (up to 26 years) by accounting for right censoring. The linearity assumption of the LRC score (considered as a continuous variable) was graphically assessed using Martingale residuals before adding family-specific random effects. We did not find evidence for a non-linear effect of the LRC score.

Ethical regulations

Leiden Longevity Study: In accordance with the Declaration of Helsinki, we obtained informed consent from all participants prior to their study entry. Good clinical practice guidelines were maintained. The study protocol was approved by the ethical committee of the Leiden University Medical Center before the start of the study (P01.113).

SEDD: The SEDD has approval for research from Regionala etikprövningsnämnden, Lund, (dnr 161/2006, dnr 627/2010), and instructions from Datainspektionen, Stockholm (dnr 1999-2005).

Reporting summary

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

Data availability

The individual-level data from the SEDD, the Statistics Sweden, and LLS are protected by Swedish and Dutch personal integrity laws, and other (privacy) regulations. As such, restrictions apply to the availability of the LLS and SEDD data, which were used under license for the current study, and so are not publicly available. All summary statistics and data, underlying the main figures, are available in the source data file. For both datasets, additional summary statistics are available upon request to Niels van den Berg (corresponding author; n.m.a.van_den_berg@lumc.nl). The LLS data is available for replication purposes upon request to P. Eline Slagboom (p.slagboom@lumc.nl) and if replication is conducted within the secure Leiden University Medical Center network environment. Researchers can gain access to the SEDD data as used in this study if relevant permissions have been obtained in accordance with the restrictions stated by the Regional Ethical Review Board, the Swedish Data Inspection Board, and Lund University (ingrid.van_dijk@ekh.lu.se). Initial responses for both LLS and SEDD will be within one week. Research requests for commercial use will not be considered. Source data are provided with this paper.

Code availability

The scripts containing the code for data pre-processing, data analyses, and simulations performed in response to the reviewers can be freely downloaded at: https://git.lumc.nl/publications/longevity-family-diseases.

References

1. Oeppen, J. & Vaupel, J. W. Broken limits to life expectancy. *Science* **296**, 1029–1031 (2002).
2. Wang, H. et al. Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1160–1203 (2020).
3. Barnett, K. et al. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* **380**, 37–43 (2012).
4. Andersen, S. L., Sebastiani, P., Dworkis, D. A., Feldman, L. & Perls, T. T. Health span approximates life span among many super-centenarians: compression of morbidity at the approximate limit of life span. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **67A**, 395–405 (2012).
5. Ash, A. S. et al. Are members of long-lived families healthier than their equally long-lived peers? Evidence from the long life family study. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 971–976 (2015).
6. Lipton, R. B. et al. Exceptional parental longevity associated with lower risk of alzheimer's disease and memory decline. *J. Am. Geriatr. Soc.* **58**, 1043–1049 (2010).
7. Galvin, A., Ukraintseva, S., Arbeev, K., Feitosa, M. & Christensen, K. Physical robustness and resilience among long-lived female siblings: a comparison with sporadic long-livers. *Aging* **12**, 15157–15168 (2020).
8. Christensen, K. et al. Mechanisms underlying familial aggregation of exceptional health and survival: a three-generation cohort study. *Aging Cell* **19**, 1–11 (2020).
9. Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional longevity does not result in excessive levels of disability. *Proc. Natl. Acad. Sci. USA* **105**, 13274–13279 (2008).
10. Ailshire, J. A., Beltran-Sanchez, H. & Crimmins, E. M. Becoming centenarians: disease and functioning trajectories of older U.S. adults as they survive to 100. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 193–201 (2015).
11. Adams, E. R., Nolan, V. G., Andersen, S. L., Perls, T. T. & Terry, D. F. Centenarian offspring: start healthier and stay healthier. *J. Am. Geriatr. Soc.* **56**, 2089–2092 (2008).
12. Gellert, P. et al. Centenarians differ in their comorbidity trends during the 6 years before death compared to individuals who died in their 80s or 90s. *J. Gerontol. Ser. A* **73**, 1357–1362 (2018).
13. Ismail, K. et al. Compression of morbidity is observed across cohorts with exceptional longevity. *J. Am. Geriatr. Soc.* **64**, 1583–1591 (2016).
14. Gueresi, P. et al. Does the longevity of one or both parents influence the health status of their offspring? *Exp. Gerontol.* **48**, 395–400 (2013).
15. Dutta, A. et al. Longer lived parents: protective associations with cancer incidence and overall mortality. *J. Gerontol. - Ser. A Biol. Sci. Med. Sci.* **68**, 1409–1418 (2013).
16. Gubbi, S. et al. Effect of exceptional parental longevity and lifestyle factors on prevalence of cardiovascular disease in offspring. *Am. J. Cardiol.* **120**, 2170–2175 (2017).
17. Evert, J., Lawler, E., Bogan, H. & Perls, T. Morbidity profiles of centenarians: survivors, delayers, and escapers. *J. Gerontol. - Med. Sci.* **58**, 232–237 (2003).
18. Zeng, Y. et al. Health consequences of familial longevity influence among the Chinese elderly. *J. Gerontol. - Ser. A Biol. Sci. Med. Sci.* **68**, 473–482 (2013).
19. Richmond, R. L., Law, J. & KayLambkin, F. Morbidity profiles and lifetime health of Australian centenarians. *Australas. J. Ageing* **31**, 227–232 (2012).
20. van den Berg, N. Family matters in unraveling human longevity. *Aging* **12**, 1–2 (2020).
21. van den Berg, N. et al. Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nat. Commun.* **10**, 35 (2019).
22. Berg, N. et al. Longevity Relatives Count score identifies heritable longevity carriers and suggests case improvement in genetic studies. *Aging Cell* **19**, 1–15 (2020).
23. van den Berg, N. et al. Longevity around the turn of the 20th century: life-long sustained survival advantage for parents of today's nonagenarians. *J. Gerontol. Ser. A* **73**, 1295–1302 (2018).
24. Mourits, R. J. et al. Intergenerational transmission of longevity is not affected by other familial factors: evidence from 16,905 Dutch families from Zeeland, 1812-1962. *Hist. Fam.* **25**, 484–526 (2020).

25. Deelen, J. et al. A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat. Commun.* **10**, 3346 (2019).

26. Pedersen, J. K. et al. The survival of spouses marrying into longevity-enriched families. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **72**, 109–114 (2017).

27. Gavrilova, N. S. & Gavrilov, L. A. Protective effects of familial longevity decrease with age and become negligible for centenarians. *J. Gerontol. Ser. A* **77**, 736–743 (2022).

28. Lichtenstein, P. et al. Environmental and heritable factors in the causation of cancer — analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85 (2000).

29. Prins, B. P., Lagou, V., Asselbergs, F. W., Snieder, H. & Fu, J. Genetics of coronary artery disease: genome-wide association studies and beyond. *Atherosclerosis* **225**, 1–10 (2012).

30. FEINLEIB, M. et al. The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results. *Am. J. Epidemiol.* **106**, 284–295 (1977).

31. Wienke, A., Holm, N. V., Skytthe, A. & Yashin, A. I. The heritability of mortality due to heart diseases: a correlated frailty model applied to danish twins. *Twin Res.* **4**, 266–274 (2001).

32. Doris, P. A. The genetics of blood pressure and hypertension: the role of rare variation. *Cardiovasc. Ther.* **29**, 37–45 (2011).

33. Zdravkovic, S. et al. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J. Intern. Med.* **252**, 247–254 (2002).

34. Drobni, Z. D. et al. Heritability of coronary artery disease: insights from a classical twin study. *Circ. Cardiovasc. Imaging* **15**, 133–141 (2022).

35. Levy, D. et al. Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* **41**, 677–687 (2009).

36. Khera, A. V. & Kathiresan, S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat. Rev. Genet.* **18**, 331–344 (2017).

37. Bevan, S. et al. Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genome-wide associations. *Stroke* **43**, 3161–3167 (2012).

38. Muñoz, M. et al. Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat. Genet.* **48**, 980–983 (2016).

39. Willemsen, G. et al. The concordance and heritability of type 2 diabetes in 34,166 twin pairs from international twin registers: The Discordant Twin (DISCOTWIN) Consortium. *Twin Res. Hum. Genet.* **18**, 762–771 (2015).

40. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).

41. Tindale, L. C., Salema, D. & Brooks-Wilson, A. R. 10-year follow-up of the Super-Seniors Study: compression of morbidity and genetic factors. *BMC Geriatr.* **19**, 58 (2019).

42. Terry, D. F., Wilcox, M., McCormick, M. A., Lawler, E. & Perls, T. T. Cardiovascular advantages among the offspring of centenarians. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **58**, M425–M431 (2003).

43. Terry, D. F., Wilcox, M. A., McCormick, M. A. & Perls, T. T. Cardiovascular disease delay in centenarian offspring. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **59**, M385–M389 (2004).

44. Terry, D. F. et al. Lower all-cause, cardiovascular, and cancer mortality in centenarians' offspring. *J. Am. Geriatr. Soc.* **52**, 2074–2076 (2004).

45. Sebastiani, P. et al. Families enriched for exceptional longevity also have increased health-span: findings from the long life family study. *Front. Public Health* **1**, 1–9 (2013).

46. Beekman, M. et al. Genome-wide association study (GWAS)-identified disease risk alleles do not compromise human longevity. *Proc. Natl. Acad. Sci. USA* **107**, 18046–18049 (2010).

47. Surendran, P. et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).

48. Bertram, L. & Tanzi, R. E. Alzheimer disease risk genes: 29 and counting. *Nat. Rev. Neurol.* **15**, 191–192 (2019).

49. Kaplanis, J. et al. Quantitative analysis of population-scale family trees with millions of relatives. *Science* **360**, 171–175 (2018).

50. Erlich, Y., Shor, T., Pe'er, I. & Carmi, S. Identity inference of genomic data using long-range familial searches. *Science* **362**, 690–694 (2018).

51. Crimmins, E. M. Social hallmarks of aging: Suggestions for geroscience research. *Ageing Res. Rev.* **63**, 101136 (2020).

52. Soininen, P. et al. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* **134**, 1781 (2009).

53. Würtz, P. et al. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -omic technologies. *Am. J. Epidemiol.* **186**, 1084–1096 (2017).

54. Van Der Meulen, A. *Life Tables And Survival Analysis* (Statistics Netherlands, The Hague, Netherlands, 2012).

55. Carolina, T., Uijvenhoven, L. & van der Laan, J. *Overlevingstafels en Longitudinale Analyse. CBS* 1–25 (Statistics Netherlands, The Hague, Netherlands, 2009).

56. van den Berg, N., Beekman, M., Smith, K. R., Janssens, A. & Slagboom, P. E. Historical demography and longevity genetics: back to the future. *Ageing Res. Rev.* **38**, 28–39 (2017).

57. Sebastiani, P., Nussbaum, L., Andersen, S. L., Black, M. J. & Perls, T. T. Increasing sibling relative risk of survival to older and older ages and the importance of precise definitions of "aging," "life span," and "longevity". *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **71**, 340–346 (2016).

58. Bizzarri, D., Reinders, M. J. T., Beekman, M., Slagboom, P. E. & van den Akker, E. B. MiMIR: R-shiny application to infer risk factors and endpoints from Nightingale Health's 1H-NMR metabolomics data. *Bioinformatics* **38**, 3847–3849 (2022).

59. R Core Team. *R: A Language And Environment For Statistical Computing* (2016).

Acknowledgements

Author contributions

N.v.d.B. is the study investigator and was responsible for initiating the study, data management, data analyses, writing the manuscript, and finalizing it. I.v.D. was responsible for the data organization and analyses of the Swedish data. M.R.-G. provided overall support on statistical analyses. P.S. and M.B. provided overall coordination. All authors contributed to the writing process leading to this paper.

Competing interests

The authors declare no competing interests.

Additional information