nature human behaviour

LETTERS

https://doi.org/10.1038/s41562-019-0562-1

Genetics and the geography of health, behaviour and attainment

Daniel W. Belsky[1,2*], Avshalom Caspi[3,4,5,6], Louise Arseneault[6], David L. Corcoran[5], Benjamin W. Domingue[7], Kathleen Mullan Harris[8], Renate M. Houts[3], Jonathan S. Mill[9], Terrie E. Moffitt[3,4,5,6], Joseph Prinz[5], Karen Sugden[3], Jasmin Wertz[3], Benjamin Williams[3] and Candice L. Odgers[10,11*]

Young people's life chances can be predicted by characteristics of their neighbourhood[1]. Children growing up in disadvantaged neighbourhoods exhibit worse physical and mental health and suffer poorer educational and economic outcomes than children growing up in advantaged neighbourhoods. Increasing recognition that aspects of social inequalities tend, in fact, to be geographical inequalities[2–5] is stimulating research and focusing policy interest on the role of place in shaping health, behaviour and social outcomes. Where neighbourhood effects are causal, neighbourhood-level interventions can be effective. Where neighbourhood effects reflect selection of families with different characteristics into different neighbourhoods, interventions should instead target families or individuals directly. To test how selection may affect different neighbourhood-linked problems, we linked neighbourhood data with genetic, health and social outcome data for >7,000 European-descent UK and US young people in the E-Risk and Add Health studies. We tested selection/concentration of genetic risks for obesity, schizophrenia, teen pregnancy and poor educational outcomes in high-risk neighbourhoods, including genetic analysis of neighbourhood mobility. Findings argue against genetic selection/concentration as an explanation for neighbourhood gradients in obesity and mental health problems. By contrast, modest genetic selection/concentration was evident for teen pregnancy and poor educational outcomes, suggesting that neighbourhood effects for these outcomes should be interpreted with care.

A challenge in understanding how neighborhoods impact people's lives is distinguishing causal effects of neighbourhood features from processes of selection in which individuals with different characteristics come to live in different neighbourhoods[6,7]. There is growing evidence that at least some neighbourhood effects are causal; in a natural experiment arising from immigration policy in Sweden and in a randomized trial of a housing voucher programme in the United States, people assigned to better-off neighbourhoods tended to have some better health outcomes[8,9]. Economic benefits of neighbourhood interventions are less clear, but may be present for children whose neighbourhoods are changed relatively early in life[10,11]. But selection effects are also apparent. For example, in one study of hurricane survivors, those in poorer health before the disaster tended to relocate to higher-poverty communities in its aftermath[12]. Selection and causation in neighbourhood effects are not mutually exclusive: both can occur[13]. Better understanding of how selection may contribute to apparent neighbourhood effects is needed to guide intervention design and policy. Where selection can be ruled out as an explanation of neighbourhood effects, neighbourhood-level interventions could be prioritized. In instances in which apparent neighbourhood effects reflect selection processes, interventions delivered to individuals or families directly might prove more effective.

To evaluate the size and scope of selection effects in neighbourhood research, methods are needed that quantify selection factors and that are not influenced by neighbourhood conditions. The ideal approach is to compare fixed characteristics between children growing up in high-risk neighbourhoods and peers growing up in better-off neighbourhoods. Because neighbourhoods may affect individuals as early as the very beginnings of their lives[3,14], traditional social science measurements are problematic. Recent discoveries from genome-wide association studies (GWAS) provide a new opportunity to quantify selection effects at the level of the individual: polygenic scores. DNA sequence is fixed at conception and is never altered by neighbourhood environments. Because children inherit their DNA sequence from their parents, measures of genetic risk form a conceptual link between familial characteristics, such as parental education, that may influence selection into neighbourhoods and children's health and social outcomes. In this article, we report proof-of-concept polygenic score analysis to quantify genetic selection into neighbourhoods.

We analysed polygenic scores and neighbourhood conditions in 1,999 young people from the Environmental Risk Longitudinal Study (E-Risk), a birth cohort ascertained from a birth registry in England and Wales and followed prospectively through 18 years of age. We studied phenotypes that represent substantial public health and economic burdens, have been linked with neighbourhood risk in previous studies, are prevalent among 18-year olds in England and Wales, and have been subject to large-scale GWAS meta-analyses:

[1]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA. [2]Robert N. Butler Columbia Aging Center, Columbia University, New York, NY, USA. [3]Department of Psychology and Neuroscience, Duke University, Durham, NC, USA. [4]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, USA. [5]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. [6]MRC Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [7]Stanford Graduate School of Education, Stanford University, Palo Alto, CA, USA. [8]Carolina Population Center and Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [9]Complex Disease Epigenetics Group, University of Exeter Medical School, Exeter, UK. [10]Department of Psychological Science, University of California at Irvine, Irvine, CA, USA. [11]Sanford School of Public Policy, Duke University, Durham, NC, USA. *e-mail: db3275@cumc.columbia.edu; codgers@uci.edu

obesity, mental health problems, teen pregnancy and poor educational outcomes. We measured children's genetic risk using four polygenic scores computed based on results from published GWAS of obesity, schizophrenia, age at first birth and educational attainment[15–18]. We measured their neighbourhoods using administrative, survey and systematic social observation[19] data collected during their childhoods. We tested for the expected associations of polygenic and neighbourhood risk with E-Risk children's development of obesity and mental health problems, teen pregnancy, earning poor educational qualifications and not being in education, employment or training (NEET), as measured during home visits at 18 years of age. To test for genetic selection effects, we tested for gene–environment correlations, in which young people who carried elevated burdens of polygenic risk tended to have grown up in more-disadvantaged neighbourhoods. To test whether gene–environment correlations reflected the passive inheritance of genetics and neighbourhood conditions from parents, we also analysed the genetics of the children's mothers. Finally, to test how genetics might become correlated with neighbourhood conditions, we tested genetic associations with neighbourhood mobility using data from 5,325 participants in the US-based National Longitudinal Study of Adolescent to Adult Health (Add Health), a nationally representative longitudinal study of American adolescents followed prospectively through their late 20s or early 30s.

As anticipated by the genetics literature, E-Risk children with higher genetic risk had more health and social problems by 18 years of age. We computed polygenic scores from published GWAS results for obesity, schizophrenia, age-at-first birth and educational attainment[15–18] using the methods described by Dudbridge[20]. This method proceeds as follows: first, single-nucleotide polymorphisms (SNPs) in the E-Risk database were matched with SNPs reported in the GWAS publications. Second, for each matched SNP, a weight is calculated equal to the number of phenotype-associated alleles multiplied by the effect size estimated in the GWAS. Finally, the average weight across all SNPs in a study member's genome is calculated to compute their polygenic score. Scores were transformed to have a cohort-wide mean = 0 and standard deviation (s.d.) = 1 for analysis.

Eighteen-year olds with a higher polygenic risk for obesity were at increased risk for obesity (relative risk (RR) = 1.26, 95% CI: 1.14–1.38); those with a higher polygenic risk for schizophrenia were at increased risk for mental health problems (RR = 1.13, 95% CI: 1.02–1.26); those with a higher polygenic risk of young age at first birth were at increased risk for teen pregnancy (RR = 1.40, 95% CI: 1.19–1.63); and those with a higher polygenic risk for low educational attainment were at increased risk for poor educational qualifications (RR = 1.46, 95% CI: 1.34–1.59) and becoming NEET (RR = 1.32, 95% CI: 1.15–1.51) (Fig. 1, and Supplementary Table 1, panel A).

An advantage of using genetics to test for potential selection effects is that genotypes cannot be caused by the neighbourhoods where children live, ruling out reverse causation. A second advantage is that genetics may provide new information over and above what can be measured from children's families[21–23]. To evaluate whether the polygenic scores that we studied provided new information over and above family history risk information, we repeated our polygenic score analysis, adding covariate adjustment for family history measures. After covariate adjustment for family history, associations of young people's polygenic scores with their health and social problems were modestly attenuated. Family history analysis is reported in Supplementary Table 1, panels B and C.

As anticipated from the neighbourhood-effects literature, children growing up in more-disadvantaged neighbourhoods were at increased risk for health and social problems by 18 years of age. Because there is no single standard to quantify neighbourhood risks[6,24], we used two different approaches to measure E-Risk children's exposure to neighbourhood disadvantage.

The first approach characterized the neighbourhoods as they are seen by businesses and the public sector, using a consumer classification system called ACORN ('A Classification of Residential Neighbourhoods'). We computed the average ACORN classification across children's home addresses when they were aged 5, 7, 10 and 12 years. According to ACORN, 22% of E-Risk cohort children grew up in 'wealthy achiever' neighbourhoods, 33% grew up in 'urban prosperity/comfortably off' neighbourhoods, 19% grew up in 'moderate means' neighbourhoods, and 26% grew up in 'hard pressed' neighbourhoods. This distribution matched the overall distributions for the United Kingdom. As an example, ACORN distributions for E-Risk families at the time of the age-12 assessment are compared to the national distribution in Fig. 2a.

The second approach characterized neighbourhoods as they are seen by social scientists and public health researchers. Ecological-risk measures were constructed from (1) geodemographic data from local governments, (2) official crime data accessed as part of an open data sharing effort about crime and policing in England and Wales, (3) a Google Street View Virtual systematic social observation[19], and (4) data from surveys of neighbourhood residents (Fig. 2b). We used these data to score neighbourhoods on their economic deprivation, physical dilapidation, social disconnection and dangerousness. We standardized scores to have mean = 50 and s.d. = 10 (T scores). We summed these four ecological-risk measures to compute one composite ecological-risk index (mean = 198, s.d. = 33) (Supplementary Fig. 1). Ecological-risk measures were correlated with ACORN classifications (for the ecological-risk index $r = 0.65$; correlations for all measures are reported in Supplementary Table 2 and Supplementary Fig. 1).

Eighteen-year olds who grew up in neighbourhoods with more disadvantaged ACORN classifications or with higher scores on the ecological-risk index were at increased risk for obesity, mental health problems, teen pregnancy, poor educational qualifications and NEET status (obesity ACORN RR = 1.20 (95% CI: 1.10–1.31), ecological-risk index RR = 1.15 (95% CI: 1.03–1.29); mental health problems ACORN RR = 1.19 (95% CI: 1.08–1.31), ecological-risk index RR = 1.30 (95% CI: 1.14–1.47); teen pregnancy ACORN RR = 1.56 (95% CI: 1.34–1.83), ecological-risk index RR = 1.55 (95% CI: 1.30–1.85); poor educational qualifications ACORN RR = 1.53 (95% CI: 1.40–1.67), ecological-risk index RR = 1.47 (95% CI: 1.33–1.62); NEET ACORN RR = 1.52 (95% CI: 1.33–1.74), ecological-risk index RR = 1.59 (95% CI: 1.36–1.85)). Figure 3a plots the risk for each health and social problem by childhood neighbourhood disadvantage. Results for all neighbourhood measures are reported in Supplementary Table 3.

We found little evidence that genetic selection/concentration explained neighbourhood risk for obesity or mental health problems. Although children's genetic risk and their neighbourhood disadvantage separately predicted their increased risk of obesity and mental health problems, polygenic risks for obesity and schizophrenia were not consistently related to neighbourhood disadvantage. Figure 3 shows this result graphically. Whereas the blue slopes in Fig. 3a document positive associations between neighbourhood disadvantage and risk for obesity and mental health problems, the red slopes in Fig. 3b reveal null associations between neighbourhood disadvantage and polygenic risk for obesity and null or weak associations between neighbourhood disadvantage and polygenic risk for schizophrenia. In the E-Risk cohort, children raised in disadvantaged neighbourhoods more often became obese by 18 years of age; however, we found no evidence for concentration of children with high polygenic risk in disadvantaged neighbourhoods (ACORN $r = -0.01$ (95% CI: −0.07 to 0.04); ecological-risk index $r = -0.01$ (95% CI: −0.08 to 0.07)). Results were similar for analysis of genetic risk for schizophrenia, although the association between children's polygenic scores and their neighbourhood ecological-risk index was significant at the $\alpha = 0.05$ level (ACORN $r = 0.04$ (95% CI:−0.01

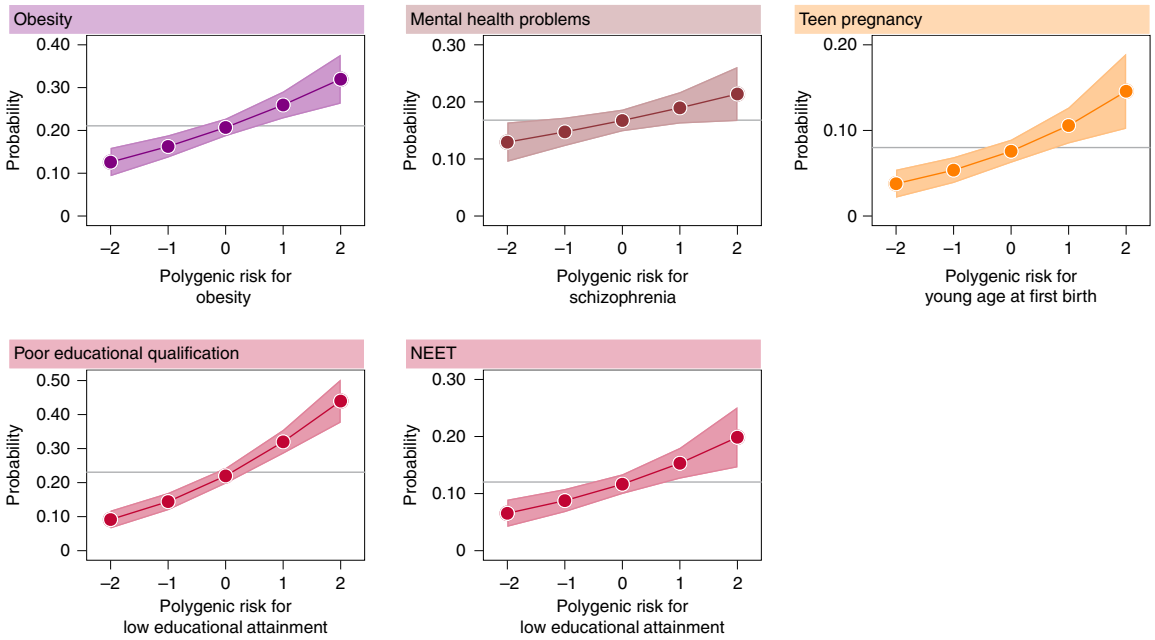

Fig. 1 | Children with a higher genetic risk had more health and social problems by 18 years of age. Graphs show the fitted probabilities of each health and social problem across the distribution of polygenic risk. Models were adjusted for sex. The grey line intersecting the *y* axis shows the frequency of the health or social problem in E-Risk. The shaded areas around the fitted slopes show 95% CIs. The probability of obesity is graphed against polygenic risk for obesity (RR = 1.26, 95% CI: 1.14–1.38, $n = 1{,}837$). The probability of mental health problems is graphed against polygenic risk for schizophrenia (RR = 1.13, 95% CI: 1.02–1.26, $n = 1{,}863$). The probability of teen pregnancy is graphed against polygenic risk for young age at first birth (RR = 1.40, 95% CI: 1.19–1.64, $n = 1{,}825$). The probabilities of poor educational qualification and NEET status are graphed against polygenic risk for low educational attainment (poor educational qualification RR = 1.47, 95% CI: 1.34–1.60, $n = 1{,}860$; NEET RR = 1.32, 95% CI: 1.15–1.52, $n = 1{,}863$) (Supplementary Table 1). Effect sizes are reported for a 1-s.d. increase in polygenic risk.

to 0.10); ecological-risk index $r = 0.08$ (95 % CI: 0.01–0.15)). Results for all neighbourhood measures are reported in Supplementary Table 4. These findings argue against neighbourhood selection/composition as a source of neighbourhood gradients in obesity and mental health problems and encourage more research to unravel the possible causal effects of neighbourhood conditions on physical and mental health.

We found evidence of genetic selection/concentration in disadvantaged neighbourhoods of children at high polygenic risk of teen pregnancy, poor educational attainment and NEET status. We tested whether children at higher polygenic risk for young age at first birth and poor educational attainment tended to grow up in more-disadvantaged neighbourhoods. They did, as measured by both ACORN classification and the composite ecological-risk index. Figure 3 shows this result graphically. The blue slopes in Fig. 3a document positive associations between neighbourhood disadvantage and risk for teen pregnancy, poor educational outcomes and NEET status. In parallel, the red slopes in Fig. 3b reveal positive associations between neighbourhood risk and polygenic risk for young age at first birth (reversed values of the age-at-first-birth polygenic score, *y* axis of the right graph; ACORN $r = 0.12$ (95% CI: 0.06–0.17); ecological-risk index $r = 0.12$ (95% CI: 0.04–0.19)) and low educational attainment (reversed values of the education polygenic score, *y* axis of the right graph; ACORN $r = 0.18$ (95% CI: 0.12–0.23); ecological-risk index $r = 0.17$ (95% CI: 0.09–0.25)). Results for all neighbourhood measures are reported in Supplementary Table 4. These findings suggest that neighbourhood selection/composition may be relevant to neighbourhood–teen pregnancy and neighbourhood–achievement gradients and encourage research to understand selection processes.

Children inherited genetic and neighbourhood risks from their parents. E-Risk children were 5–12 years of age during the period when neighbourhood data were collected. It is unlikely that they actively selected themselves into different types of neighbourhoods. Instead, a hypothesis for why children's polygenic and neighbourhood risks are correlated is that both risks are inherited from their parents. According to this hypothesis, genetics influence parents' characteristics and behaviours, which in turn affect where they live. Children subsequently inherit their parents' genetics and their neighbourhoods. As an initial test of this hypothesis, we analysed genetic data that we collected in E-Risk from children's mothers ($n = 858$ with children included in analysis). (E-Risk did not collect fathers' DNA.) As expected, polygenic scores were correlated between E-Risk participants and their mothers ($r = 0.50$–0.52). We first tested whether mothers' polygenic scores were associated with neighbourhood disadvantage. Parallel to the results from analysis of children's genetics, we did not detect associations of mothers' obesity and schizophrenia polygenic scores with neighbourhood disadvantage ($r = 0.00$–0.04). Also consistent with analysis of children, mother's age-at-first-birth and educational attainment polygenic scores were associated with neighbourhood disadvantage (effect sizes $r = 0.14$–0.21; Supplementary Table 5). Next, we repeated the analysis of association between children's polygenic scores and neighbourhood disadvantage, this time including a covariate for the mother's polygenic score. Consistent with the hypothesis that children's polygenic and neighbourhood risks are correlated because both risks are inherited from their parents, covariate adjustment for mothers' polygenic scores reduced the magnitudes of associations between children's polygenic scores and their neighbourhood disadvantage by more than half (Supplementary Table 6 and Supplementary Fig. 2).

Polygenic risk for teen pregnancy and low educational attainment predicted downward neighbourhood mobility among participants in the US National Longitudinal Study of Adolescent to Adult

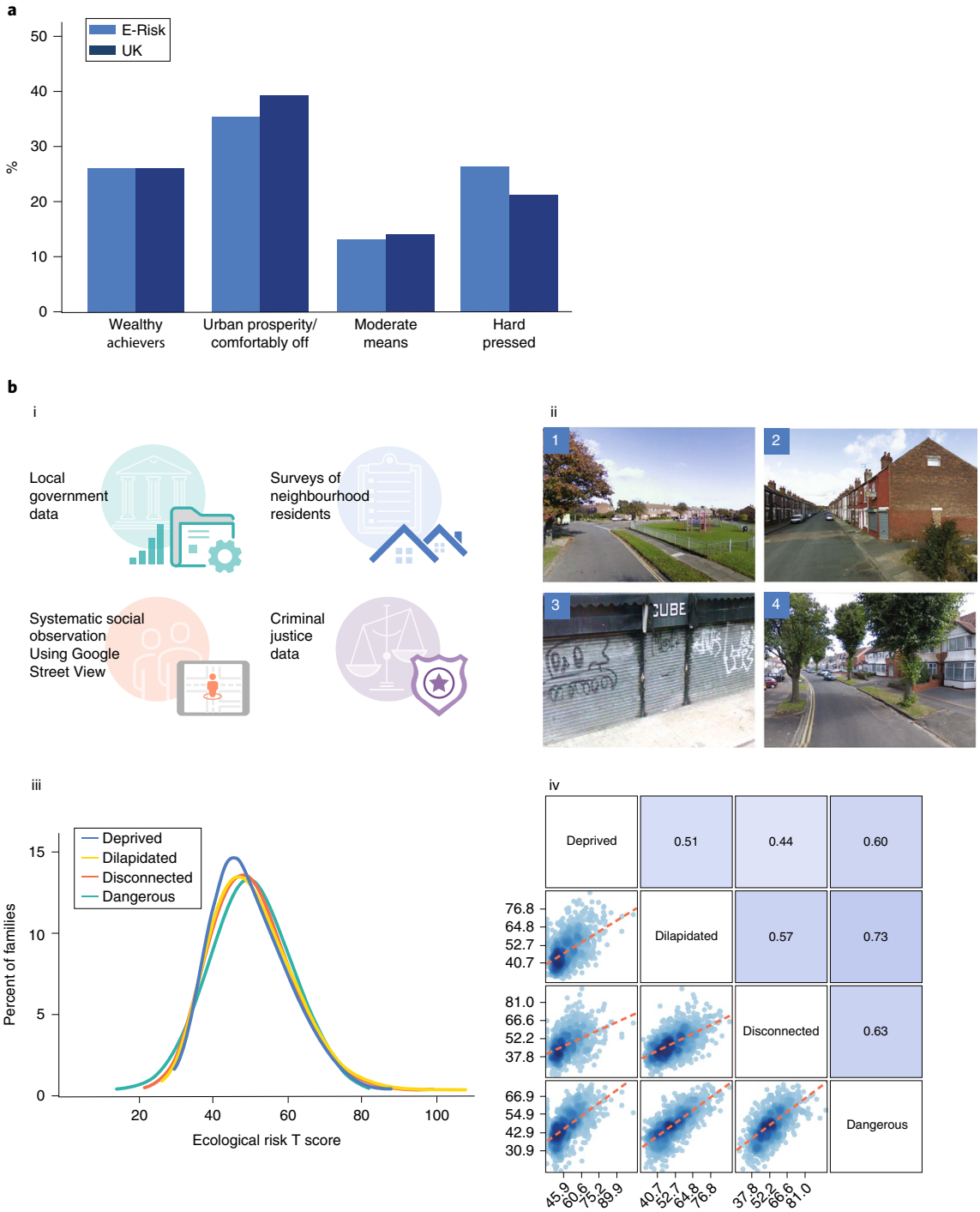

Fig. 2 | Quantification of E-Risk families' neighbourhood disadvantage using ACORN and a composite ecological-risk index. a, Distributions of ACORN classifications for E-Risk families at the time of the age-12 interview ($n = 1{,}008$ families with genetic data) and the corresponding distribution for the United Kingdom obtained from http://doc.ukdataservice.ac.uk/doc/6069/mrdoc/pdf/6069_acorn_userguide.pdf. **b**, The figure contains four cells (i–iv). (i) The four sources of data used for ecological-risk assessment: geodemographic data from local governments, resident surveys, Google Street View systematic social observation and official crime data. Image created by Motsavage Design/Bigstock.com. (ii) Images illustrating: a well-kept neighbourhood, visible play area for children, and roads and sidewalks in good condition (1); evidence of graffiti, poorly kept sidewalk and trash container, and sidewalks in fair condition (2); deprived residential area, vacant lot in poor condition, a heavy amount of litter and sidewalks in poor condition (3); and comfortably off residential area, roads and sidewalks in good conditions, and no signs of litter, graffiti or other signs of disorder (4). Images are from Google Street View. (iii) Distributions of four ecological-risk measures derived from these data: economic deprivation, physical dilapidation, social disconnectedness and danger. The values of the ecological-risk measures are expressed as T scores (mean = 50, s.d. = 10) ($n = 987$ families with genetic data). (iv) A matrix of the ecological-risk measures, illustrating their correlation with one another (see Supplementary Table 2). The matrix cells below and to the left of the measures show scatter plots of their association. The matrix cells above and to the right of the measures show their correlation expressed as Pearson's r ($n = 973$ families with genetic data and data on all four ecological-risk measures). The dashed red regression lines illustrate the slopes of associations between each pair of measures.

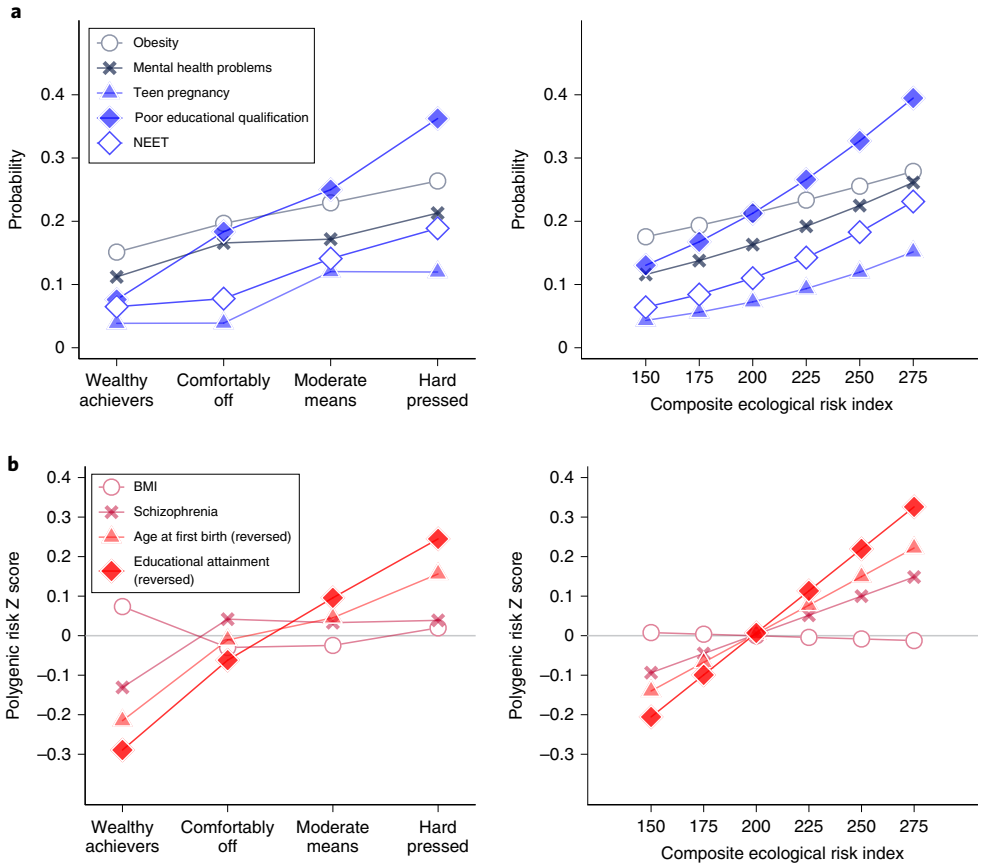

Fig. 3 | Neighbourhood gradients in obesity, mental health problems, teen pregnancy, poor educational qualifications, NEET status and in the genetic risk for these phenotypes. a, The neighbourhood risk gradient for each health and social problem. The *y* axis shows the probability of having a given problem at varying levels of neighbourhood risk. Left: probabilities by ACORN classification ($n=1,857$). Right: the predicted probabilities for a series of values of the composite ecological-risk index ($n=1,822$). Effect sizes associated with a one-category increase in disadvantage defined by ACORN and with a 1-s.d. increase in disadvantage defined by the ecological-risk index, respectively, were: obesity (RR = 1.20 (95% CI: 1.10–1.31); RR = 1.15 (95% CI: 1.03–1.29)); mental health problems (RR = 1.19 (95% CI: 1.08–1.31); RR = 1.30 (95% CI: 1.14–1.47)); having a teen pregnancy (RR = 1.56 (95% CI: 1.34–1.83); RR = 1.55 (95% CI: 1.30–1.85)); poor educational qualifications (RR = 1.53 (95% CI: 1.40–1.67); RR = 1.47 (95% CI: 1.33–1.62)) and NEET (RR = 1.52 (95% CI: 1.33–1.74); RR = 1.59 (95% CI: 1.36–1.85)) (Supplementary Table 3). **b**, The neighbourhood risk gradient for each polygenic score. The *y* axis shows the polygenic risk on a *z*-score scale (mean = 0, s.d. = 1) at varying levels of neighbourhood risk; gray lines indicate a *z*-score value of 0. Left: polygenic risk by ACORN classification ($n=1,441$). Right: polygenic risk for a series of values of the composite ecological-risk index ($n=1,414$). Effect sizes with disadvantage defined by ACORN and by the ecological-risk index, respectively, were: body mass index (BMI) polygenic score ($r=-0.01$ (95% CI: −0.07 to 0.04); $r=-0.01$ (95% CI: −0.08 to 0.07)); schizophrenia polygenic score ($r=0.04$ (95% CI: −0.01 to 0.10); $r=0.08$ (95% CI: 0.01–0.15)); age-at-first-birth polygenic score ($r=0.12$ (95% CI: 0.06–0.17); $r=0.12$ (95% CI: 0.04–0.19)); and educational attainment polygenic score ($r=0.18$ (95% CI: 0.12–0.23); $r=0.17$ (95% CI: 0.09–0.25)) (Supplementary Table 4). Sample sizes in **b** are smaller than sample sizes in **a** because the polygenic score analysis shown in **b** included only one member of each monozygotic twin pair.

Health (Add Health). If children's genetic and neighbourhood risks are correlated because they inherit both risks from their parents, the next question is how parents' genetics come to be correlated with neighbourhood risks. A hypothesis is that parents' genetics influence their characteristics and behaviour in ways that affect where they are able to live. To test this hypothesis, data are needed that observe the neighbourhood mobility process in which people leave the homes where they grew up and selectively end up in new neighbourhoods. Because E-Risk began collecting information on children's mothers only after the children were born, data were not collected on the mothers' own childhood neighbourhoods. Thus, to test how polygenic risks might influence patterns of neighbourhood mobility, we turned to a second data set: Add Health. Add Health first surveyed participants when they were secondary school students living with their parents. Add Health has since followed participants into their late 20s and early 30s[25], when most were living in new neighbourhoods ($n=5,325$ with genetic and neighbourhood data; 86% lived >1 km from the address where they was first surveyed). We used the Add Health genetic database and neighbourhood measures derived from US Census data to test whether the polygenic risk for obesity, schizophrenia, teen pregnancy and low educational attainment predicted downward neighbourhood mobility, that is, young adults coming to live in more-disadvantaged neighbourhoods relative to the neighbourhood they lived in with their parents.

Add Health participants' polygenic risks for obesity and schizophrenia showed weak or null associations with neighbourhood disadvantage when they were first surveyed in their parents' homes as secondary school students (polygenic risk for obesity $r=0.03$ (95% CI: 0.00–0.06); polygenic risk for schizophrenia $r=-0.01$ (95% CI: −0.03 to 0.02)) and when they were followed-up in their 20s and 30s (polygenic risk for obesity $r=0.04$ (95% CI: 0.01–0.07); polygenic risk for schizophrenia $r=-0.03$ (95% CI: −0.05 to 0.00)). Findings were similar in the neighbourhood mobility analysis (polygenic risk for

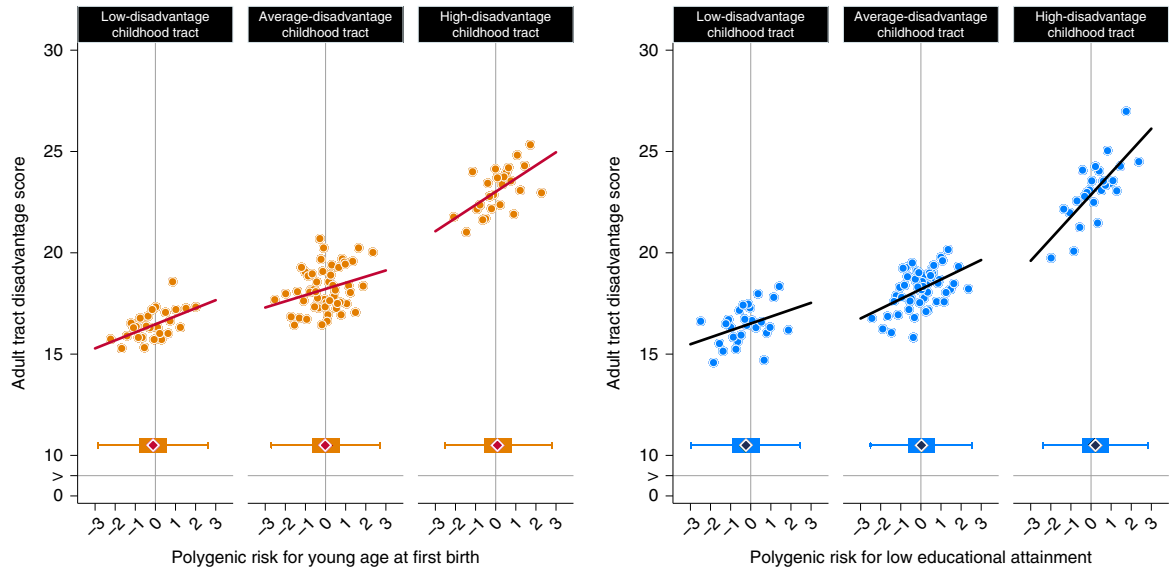

Fig. 4 | Age-at-first-birth and education polygenic score association with neighbourhood mobility in the Add Health study. The figure plots polygenic risk associations with adult neighbourhood disadvantage at the census tract level for Add Health participants who grew up in low-disadvantage, average-disadvantage and high-disadvantage census tracts (n = 5,325). For the figure, low-disadvantage, average-disadvantage and high-disadvantage census tracts were defined as the bottom quartile, middle 50% and top quartile of the childhood tract disadvantage score distribution, respectively. The individual graphs show binned scatter plots in which each plotted point reflects the average x and y coordinates for a 'bin' of 50 Add Health participants. The regression lines are plotted from the raw data. The box-and-whisker plots at the bottom of the graphs show the distribution of polygenic risk for each childhood neighbourhood disadvantage category. The diamond in the middle of the box shows the median, the box shows the interquartile range, and the whiskers show the upper and lower bounds defined by the 25th percentile minus 1.5× the interquartile range and the 75th percentile plus 1.5× the interquartile range, respectively. The vertical line intersecting the x axis shows the cohort average polygenic risk. The figure illustrates three findings. First, adult participants tended to live in census tracts with similar levels of disadvantage to the census tracts where they grew up. Second, children's polygenic risks and their neighbourhood disadvantage were correlated; the box plots show that polygenic risk tended to be lower for participants who grew up in low-disadvantage tracts and higher for participants who grew up in high-disadvantage tracts. Third, across strata of childhood neighbourhood disadvantage, children at higher polygenic risk tended to move to more-disadvantaged census tracts no matter where they grew up.

obesity r = 0.03 (95% CI: 0.00–0.05); polygenic risk for schizophrenia r = −0.02 (95% CI: −0.05 to 0.00)). These findings bolster conclusions from the E-Risk analysis that genetic selection/concentration is likely to be a trivial factor in neighbourhood gradients in obesity and mental health problems, although the obesity polygenic score association with neighbourhood risk was significant at the α = 0.05 level in Add Health and therefore did not fully replicate the findings in E-Risk.

By contrast, Add Health participants with a higher polygenic risk for teen pregnancy and low educational attainment tended to have grown up in more-disadvantaged neighbourhoods (r = 0.07 (95% CI: 0.05–0.10) for the age-at-first-birth polygenic score and r = 0.17 (95% CI: 0.14–0.19) for the educational attainment polygenic score) and to live in more-disadvantaged neighbourhoods when they were followed-up in their 20s and 30s (adult neighbourhood r = 0.09 (95% CI: 0.06–0.11) for the age-at-first-birth polygenic score and r = 0.13 (95% CI: 0.10–0.16) for the educational attainment polygenic score). In the neighbourhood mobility analysis, participants with a higher polygenic risk for teen pregnancy and low educational attainment tended to move to more-disadvantaged neighbourhoods relative to the neighbourhoods where they lived with their parents when they were first surveyed (downward mobility r = 0.06 (95% CI: 0.03–0.08) for the age-at-first-birth polygenic score and r = 0.07 (95% CI: 0.05–0.09) for the educational attainment polygenic score; Fig. 4). These findings bolster conclusions from the E-Risk analysis that genetic selection/concentration may contribute to neighbourhood gradients in teen pregnancy and poor educational outcomes, although this contribution may be small.

To put polygenic score–neighbourhood associations in context, we conducted a SNP heritability analysis to estimate the upper bound of the information that polygenic scores could contain about participants' neighbourhoods. Traditional heritability analysis compares phenotypic similarity between more and less closely related relatives, for example, monozygotic twins who share 100% of their DNA and dizygotic twins who share 50% of their DNA. Analysis of how much more similar monozygotic twins are than dizygotic twins can be used as an estimate of how much of the variation in the phenotype can be attributed to genetic differences between people. This estimate is called heritability. SNP heritability analysis operates on the same principle, but instead of using known genetic similarities of relatives, SNP heritability analysis uses measured genetic similarity derived from the same genome-wide SNP data used to construct polygenic scores[26]. An advantage of SNP heritability is that it can be estimated from samples of unrelated individuals. We estimated SNP heritability of neighbourhood disadvantage in Add Health using the GCTA software[27]. We tested whether Add Health participants who were more genetically similar to one another also tended to live in neighbourhoods with more similar levels of disadvantage. We focused GCTA analysis on the neighbourhoods where Add Health participants had moved to by the time of the 2008 follow-up, when the participants were in their 20s and 30s. Analysis included unrelated (genetic-relatedness matrix (GRM) < 0.05) participants with available neighbourhood and genotype data (n = 4,655). (We did not conduct GCTA analysis in the E-Risk sample because power calculations suggested that plausible SNP heritabilities could not be distinguished from zero in that cohort; see Methods.) The result showed that neighbourhood disadvantage was heritable; about 16% of the variance in Add Health participants' adult neighbourhood disadvantage could be explained by their genetics (95% CI:

0.01–0.32). Neighbourhood mobility was somewhat less heritable; 9% of the variance in neighbourhood mobility could be explained by genetics, but the confidence interval for the estimate overlapped zero (95% CI: −0.06 to 0.24). Comparing the results of our polygenic score analysis with this analysis of SNP heritability suggests that, in Add Health, participants' age-at-first-birth polygenic scores explain about 5% of the SNP heritability in neighbourhood disadvantage and their education polygenic scores explain about 11% of the SNP heritability in neighbourhood disadvantage. For mobility, these fractions were 4% and 6% for age-at-first-birth and education polygenic scores, respectively.

In summary, sociogenomic analyses testing the concentration of polygenic risks for health, behaviour and social problems in children growing up in disadvantaged neighbourhoods yielded three findings: first, we found little consistent evidence for the concentration of polygenic risk for obesity or polygenic risk for mental health problems in children growing up in disadvantaged neighbourhoods. Second, by contrast, we found consistent evidence for the concentration of polygenic risks for teen pregnancy and low achievement. The concentration of polygenic risks was mostly explained by children's inheritance of both neighbourhood and polygenic risks from their parents. Third, selective mobility may contribute to the concentrations of risks. In the neighbourhood mobility analysis that followed young people living with their parents during adolescence to where they lived as adults nearly two decades later, participants with a higher polygenic risk for teen pregnancy and low achievement exhibited downward neighbourhood mobility, moving to more-disadvantaged neighbourhoods across follow-up.

Large investments are being made in neighbourhood-level policies and programmes intended to improve the health and well-being of residents. These investments are based on exciting new findings demonstrating causal long-term effects of local neighbourhoods on health and possibly economic outcomes for children moving out of poverty[9–11]. The promise of place-based intervention efforts is that they can improve, at scale, the lives of residents and, for children, break the intergenerational transmission of poverty and lack of opportunity. At the same time, GWAS are revealing genetic predictors of the health outcomes, behaviours and attainments that place-based interventions seek to modify. We carried out a study of genetic selection into neighbourhoods to test how genes and geography combine.

We did not find consistent evidence of genetic selection effects into neighbourhoods for obesity and mental health problems. A previous Swedish study detected evidence of a gene–neighbourhood correlation between the schizophrenia polygenic score and a commercial database measure of neighbourhood deprivation ($r=0.04$, $n=\sim7{,}000$)[28]. The magnitude of the association was about the same as we observed for the E-Risk twins ($r=0.04$) using the commercial database ACORN measure. However, this association was not replicated in Add Health, in which the association was not significant and was in the opposite direction ($r=-0.01$ to -0.03). It is possible that selective non-participation in research related to genetic liability for schizophrenia could limit the ability to detect these associations in some data sets[29,30]. Our results nevertheless document consistent evidence (across measures and samples) of a gene–neighbourhood correlation for GWAS discoveries for age at first birth and educational attainment, and less so for GWAS discoveries for obesity and schizophrenia.

We found consistent evidence of genetic selection effects into neighbourhoods for teen pregnancy, poor educational qualifications and NEET status. These findings are consistent with recent findings in sociology about how neighbourhood residents come to be both physically and economically 'stuck in place' across generations[31,32]. Teen pregnancy and poor outcomes in education and the workplace can trap parents and their children in disadvantaged neighbourhoods, causing a clustering of individual-level and neighbourhood-level risks. This has led to calls for multigenerational and multi-level intervention efforts to break the cycle of disadvantage. Although our findings show that selection is at work for these key outcomes, the effects documented are unlikely to be large enough to fully account for neighbourhood gradients. Consistent evidence for both selection (from us) and social causation (in the larger literature) means that policies and interventions will need to target resources at both people and place to be effective.

Our findings make three contributions. First, they make a methodological and conceptual contribution by integrating genetics and social science in the rapidly developing field of social geography. We know that the places where children grow up are associated with whether they thrive. The challenge in neighbourhood research is to sort out selection from causation. Here, we take a fresh look at this classic problem using new information from genomics research. DNA sequence differences between people index differences in liability to health and social outcomes, and DNA cannot be influenced by neighbourhoods. As the price of generating genetic data continues to fall, measurements of these DNA differences can provide tools to advance social science research into the effects of place.

Second, findings shed light on how genetics and environments combine to influence children's development. Genetics contribute to the effects of place by influencing where people choose to live, are forced to live or otherwise end up living. For young people from E-risk and Add Health, some genetic risks were patterned across neighbourhoods, presumably reflecting the children's inheritance of genetics that influenced where their parents were able to live. This patterning was apparent for genetics linked to teen pregnancy and poor education, but not with genetics linked to mental health problems or obesity. One interpretation is that teen pregnancy and poor education are more proximate causes of economic circumstances that determine where one can live than, for example, obesity. Consistent with this interpretation, young people from Add Health who carried higher levels of polygenic risk for teen pregnancy and poor educational outcomes showed patterns of downward neighbourhood mobility, tending to move to worse-off neighbourhoods in young adulthood relative to the neighbourhoods where they grew up. By contrast, Add Health participants' polygenic risk for obesity and schizophrenia showed trivial or null associations with their neighbourhood mobility. Findings document that, even though the risk for highly heritable health problems such as obesity and schizophrenia may be patterned across neighbourhoods, genetic risks for these conditions may not be. More broadly, findings highlight that a phenotype being heritable does not imply that social risk factors are necessarily genetically confounded.

Third, findings provide evidence that many children are growing up subject to correlated genetic and place-related risks, particularly for teen pregnancy and attainment failure. The polygenic score–neighbourhood correlations that we observed are too small to account entirely for neighbourhood effects, but genetic and neighbourhood risks may act in combination. Thus, neighbourhood interventions can be conceptualized, in part, as breaking up gene–environment correlations, lending urgency to the development of effective place-based interventions. To this end, genetically informed designs may offer opportunities to advance intervention research. For example, comparative studies could test whether correlations between genetic and neighbourhood risks vary across cities governed by different urban planning strategies. Intervention studies could also actively incorporate genetic information: trials of neighbourhood interventions can improve the precision of their treatment-effect estimates by including polygenic score measurements as control variables to account for unmeasured differences between participants[33].

We acknowledge the limitations of our study. Foremost, our measures of genetic risk are imprecise. They explain only a fraction of the genetic variance in risk estimated from family-based genetic

models: the polygenic score for educational attainment explains >10% of phenotypic variance, polygenic scores for body mass index and schizophrenia explain 6–7% of phenotypic variance, and the age-at-first-birth polygenic score explains 1% of phenotypic variance[15–18], whereas heritabilities of these traits and behaviours estimated in family-based studies tend to be much higher[34]. As a consequence, our estimates of gene–neighbourhood correlations should be considered lower-bound estimates. Second, a related limitation is that the different polygenic scores had different amounts of power to detect associations with neighbourhood risk, with the education polygenic score having more power than the others. Nevertheless, we had more power for the body mass index and schizophrenia polygenic score analyses than we did for the age-at-first-birth polygenic score analysis, and yet genetic associations with neighbourhood risk were much larger for the age-at-first-birth polygenic score than for the body mass index and schizophrenia polygenic scores. This pattern of results held in both the E-Risk and Add Health studies. This pattern of results was also consistent with a previous analysis that used the linkage disequilibrium (LD) score regression method to test genetic correlations of an area-level measure of social deprivation with health and social outcomes[35]. Third, the magnitudes of observed gene–neighbourhood correlations in our study were small. For example, the strongest gene–neighbourhood correlations that we observed were for the educational attainment polygenic score ($r = \sim 0.17$). Based on our analysis, neighbourhood differences in this polygenic score between the highest-risk and lowest-risk neighbourhoods could account for, at most, only about 15% of the observed differences in poor educational qualifications and about 10% of the observed differences in NEET status between these neighbourhoods (Supplementary Methods). As GWAS sample sizes continue to grow, more precise measurements will become available[36]. More predictive polygenic scores could potentially strengthen measured gene–neighbourhood correlations and explain increasing fractions of neighbourhood gradients in health and social outcomes.

An additional limitation is that E-Risk data come from a single birth cohort in a single country, and thus reflect a relatively specific geographical and historical context. Findings that polygenic risk of teen pregnancy and low educational attainment were correlated with neighbourhood disadvantage did replicate in the US-based Add Health Study. Add Health neighbourhood risk was measured from tract-level US Census data describing broad social and economic conditions and is thus less geographically precise than the small-area ACORN and ecological-risk assessment data analysed in E-Risk. Thus, the Add Health analysis is not a direct replication of our E-Risk findings. Instead, the consistent results across two studies of different populations measured using different methods argue for the overall robustness of our findings.

Our analysis in both the E-Risk and Add Health studies was limited to individuals of European descent. This restriction was necessary to match the ancestry of our analytic sample with the ancestry of the samples studied in the GWAS used to calculate polygenic scores, which is the recommended approach[37]. As polygenic scores are developed for populations of non-European ancestry, replication in these populations should be a priority.

Finally and fundamentally, our results cannot establish causal relationships between genetics, neighbourhood risk and health, behaviour and attainment outcomes. It could be that genetics influence reproductive behaviour and educational attainment in ways that affect neighbourhood mobility. But there are alternative possibilities. For example, if neighbourhood risks cause teen pregnancy and school failure, GWAS of age at first birth and educational attainment could identify genetics that influence exposure to those causal neighbourhood risks. If and when large-scale GWAS of neighbourhood mobility are conducted, emerging statistical methods, such as two-sample Mendelian randomization[38,39], may be able to clarify this important causal question.

The observation of gene–neighbourhood correlations does not suggest that residents in disadvantaged neighbourhoods will not benefit from neighbourhood-level interventions. It simply means that policy-makers should not overinterpret neighbourhood effects in purely causal terms. For example, people observed to live in a friendly suburb, remote ranch, quaint village and luxury high-rise are not found in those neighbourhoods randomly by accident; people end up in such locations selectively. But regardless of the location, they all respond to incentives and opportunities. More precise quantifications of selection processes that influence where people live can help to inform policies and programmes to craft incentives and opportunities that promote healthy development for everyone.

Methods

E-Risk. *Sample.* Participants were members of the E-Risk twin study, which tracks the development of a birth cohort of 2,232 British children. The sample was drawn from a larger birth register of twins born in England and Wales in 1994–1995 (ref. [40]). Full details about the sample are reported elsewhere[41]. Briefly, the E-Risk sample was constructed in 1999–2000, when 1,116 families (93% of those eligible) with same-sex 5-year-old twins participated in home visit assessments. The sample includes 56% monozygotic and 44% dizygotic twin pairs; sex is evenly distributed within zygosity (49% male). Families were recruited to represent the population of families with newborn babies in England and Wales in the 1990s, on the basis of residential location throughout England and Wales, and the mother's age. Teenaged mothers with twins were over-selected to replace high-risk families who were selectively lost to the register through non-response. Older mothers having twins via assisted reproduction were under-selected to avoid an excess of well-educated older mothers. These strategies ensured that the study sample represents the full range of socioeconomic conditions in Great Britain[19].

Follow-up home visits were conducted when the children were 7 years of age (98% participation), 10 years of age (96% participation), 12 years of age (96% participation) and, in 2012–2014, 18 years of age (93% participation). There were no differences between those who did and did not take part at 18 years of age in terms of socioeconomic status assessed when the cohort was initially defined ($\chi^2 = 0.86$, $P = 0.65$), age-5 IQ scores ($t = 0.98$, $P = 0.33$), or age-5 internalizing or externalizing behaviour problems ($t = 0.40$, $P = 0.69$ and $t = 0.41$, $P = 0.68$, respectively). Home visits at 5, 7, 10 and 12 years of age included assessments with participants as well as their mother; the home visit at 18 years of age included interviews only with the twin participants. All interviews at the age-18 assessment were conducted after each study participant's eighteenth birthday. Each twin participant was assessed by a different interviewer. The joint Research and Development Office of South London and Maudsley and the Institute of Psychiatry Research Ethics Committee approved each phase of the study. Parents gave informed consent and twins gave assent between 5 and 12 years of age; twins gave informed consent at 18 years of age.

Genetic data. We used Illumina HumanOmni Express 12 BeadChip arrays (version 1.1; Illumina) to assay common SNP variation in the genomes of cohort members. We imputed additional SNPs using the IMPUTE2 software (version 2.3.1; https://mathgen.stats.ox.ac.uk/impute/impute_v2.html)[42] and the 1000 Genomes Phase 3 reference panel[43]. Imputation was conducted on autosomal SNPs appearing in dbSNP (version 140; http://www.ncbi.nlm.nih.gov/SNP/)[44] that were 'called' in more than 98% of the samples. Invariant SNPs were excluded. Pre-phasing and imputation were conducted using a 50 million-base pair sliding window. We analysed SNPs in Hardy–Weinberg equilibrium ($P > 0.01$). The resulting genotype databases included genotyped SNPs and SNPs imputed with 90% probability of a specific genotype among European-descent E-Risk members ($n = 1{,}999$ children in 1,011 families). The same procedure was used to construct the genetic database for study members' mothers. Genetic data were available for $n = 860$ mothers of the study members in our genetic analysis sample.

Polygenic scoring. We computed polygenic scores for obesity (body mass index), schizophrenia, age at first birth and educational attainment from published GWAS results[15–18]. We computed these polygenic scores because the GWAS on which they are based are among the largest and most comprehensive available and their target phenotypes are established as having strong geographical gradients in risk. For example, in the case of schizophrenia, there is a long-running debate about hypotheses of social causation, in which ecological risks contribute to schizophrenia pathogenesis, and social drift, in which genetic liability to schizophrenia causes downward social mobility[45,46].

Polygenic scoring was conducted following the method described by Dudbridge[20] using the PRSice software[47]. Briefly, SNPs reported in GWAS results were matched with SNPs in the E-Risk database. For each SNP, the count of phenotype-associated alleles (that is, alleles associated with a higher body mass index, increased risk of schizophrenia, younger age at first birth or less educational attainment, depending on the score being calculated) was weighted according to

the effect estimated in the GWAS. Weighted counts were averaged across SNPs to compute polygenic scores. We used all matched SNPs to compute polygenic scores irrespective of nominal significance in the GWAS.

Polygenic score analysis may be biased by population stratification, the non-random patterning of allele frequencies across ancestry groups[37,48]. To address residual population stratification within the members of European descent of the E-Risk sample, we conducted principal component analysis[49]. We computed principal components from the genome-wide SNP data with the PLINK software[50] using the command 'pca'. One member of each twin pair was selected at random from each family for this analysis. SNP loadings for principal components were applied to co-twin genetic data to compute principal component values for the full sample. We residualized polygenic scores for the first ten principal components estimated from the genome-wide SNP data[51] and standardized residuals to have mean = 0 and s.d. = 1 for analysis.

Neighbourhood disadvantage. We characterized the neighbourhoods where E-Risk study members grew up using two approaches. The first approach characterized the neighbourhoods as they are seen by businesses and the public sector, using a consumer classification system called ACORN. The second approach characterized neighbourhoods as they are seen by social scientists and public health researchers, using ecological-risk assessment methods.

Neighbourhood disadvantage measured by consumer classification. We used a geodemographic classification system, ACORN, developed as a tool for businesses interested in market segmentation by CACI (CACI, http://www.caci.co.uk/). This is a proprietary algorithm that is sold to businesses, but which CACI made available to our group for research purposes. ACORN classifications were derived from the analysis of census and consumer research databases. ACORN classifies neighbourhoods, in order of least disadvantaged to most disadvantaged, as 'wealthy achievers', 'urban prosperity', 'comfortably off', 'moderate means' or 'hard pressed'. For analysis, we combined neighbourhoods classified in the 'urban prosperity' and 'comfortably off' categories because very few children lived in 'urban prosperity' neighbourhoods. (Nationally, fewer children live in neighbourhoods characterized by 'urban prosperity'.) We obtained ACORN classifications for the output areas where E-Risk families lived[19]. Output areas are the smallest unit at which UK Census data are provided and they reflect relatively small geospatial units of about 100–125 households. Households were classified based on street address at the time of the age-5, age-7, age-10 and age-12 in-home visits. We assigned children the average neighbourhood classification across these four measurements. ACORN classifications were available for $n = 1{,}993$ children in 1,008 families in the genetic sample.

Neighbourhood disadvantage measured by ecological-risk assessment. Ecological-risk assessment was conducted by combining information from four independent sources of data: geodemographic data from local governments, official crime data from the UK police, a Google Street View-based systematic social observation, and surveys of neighbourhood residents conducted by the E-Risk investigators when the E-Risk children were 13–14 years of age. These data sources are described in detail in the Supplementary Methods.

We used these four data sources to measure the neighbourhood ecological risk in four domains: deprivation, dilapidation, disconnection and danger. Deprivation was measured with the Department of Community and Local Government Index of Multiple Deprivation. Dilapidation was measured from resident ratings of problems in their neighbourhood (for example, litter, vandalized public spaces and vacant storefronts) and independent raters' assessments of these same problems based on the 'virtual walk-through' using Google Street View. Disconnection was measured from resident surveys assessing neighbourhood collective efficacy and social connectedness. Neighbourhood collective efficacy was assessed via the resident survey using a previously validated ten-item measure of social control and social cohesion[52]. Residents were asked about the likelihood that their neighbours could be counted on to intervene in various ways if, for example: 'children were skipping school and hanging out on a street corner' or 'children were spray-painting graffiti on a local building'. They were also asked how strongly they agreed that, for example: 'people around here are willing to help their neighbours,' or 'this is a close-knit neighbourhood' (item responses: 0–4). Social connectedness was assessed based on indicators of intergenerational closure ('if any of your neighbours' children did anything that upset you would you feel that you could speak to their parents about it?'), reciprocated exchange (for example, 'would you be happy to leave your keys with a neighbour if you went away on holiday?') and friendship ties (for example, 'do you have any close friends that live in your neighbourhood') among neighbours developed in previous research[53]. Dangerousness was measured from police records of crime incidence, from neighbourhood residents' ratings of how much they feared for their safety and whether they had been victimized in their neighbourhood, and from independent raters' assessments of neighbourhood safety based on the 'virtual walk-through' using Google Street View (Fig. 2b).

For each of the four domains, we constructed a measure of ecological risk as follows. First, variables with skewed distribution were log transformed. Second, values were standardized to have mean = 50 and s.d. = 10. (For domains in which multiple resident survey or systematic social observation measures were available, we combined values within the measurement method before standardizing.) Finally, scores were averaged across measurement method within each domain. The resulting scales of deprivation, dilapidation, disconnection and danger were approximately normally distributed (Fig. 2b). Neighbourhoods' ecological-risk levels on these four measures were correlated (Pearson's $r = 0.4–0.7$; Fig. 2b). We computed the composite ecological-risk index by summing values across the four risk domains. Values were pro-rated for families with data on at least three of the four domains. Ecological-risk index values were available for $n = 1{,}954$ children in 987 families in the genetic sample.

Phenotypes. We selected phenotypes for analysis that represented substantial public health and economic burden, had been linked with neighbourhood risk in previous studies, were prevalent among 18-year olds in the United Kingdom at the time data were collected and had been subject to large-scale GWAS meta-analyses: obesity, mental health problems, teen pregnancy and poor educational outcome.

Obesity. Trained research workers took anthropometric measurements of study members when they were 18 years of age. Body mass index was computed as weight in kilograms over squared height in metres. The waist/hip ratio was calculated by dividing waist circumference by hip circumference. We defined obesity using the US Centers for Disease Control and Prevention threshold of body mass index > 30 and the World Health Organization recommendation of a waist/hip ratio of >0.90 for men and >0.85 for women[54]. Of the analysis sample, 21% met at least one of these criteria, similar to the prevalence for 16–24-year olds in the United Kingdom[55].

Mental health problems. Our measure of mental health problems is a general factor of psychopathology, the 'p-factor,' derived from confirmatory factor analysis of symptom-level psychopathology data collected at 18 years of age, when E-Risk participants were assessed in private interviews about alcohol dependence, tobacco dependence, cannabis dependence, conduct disorder, attention-deficit hyperactivity disorder, depression, generalized anxiety disorder, post-traumatic stress disorder, eating disorder and thought/psychotic disorders[56]. The 'p-factor' indexes liability to develop a wide spectrum of mental health problems[57]. We classified E-Risk study members reporting psychiatric symptoms 1 s.d. or more above the cohort norm as having mental health problems. Of the analysis sample, 17% met this criterion.

Teen pregnancy. Getting pregnant (for women) and getting someone pregnant (for men) was assessed as part of a computer-assisted interview about reproductive behaviour at the age-18 interview. Of the analysis sample, 8% (6% of men and 9% of women) reported a teen pregnancy.

Poor educational qualifications. Poor educational qualification was assessed by whether participants did not obtain or scored a low average grade (grade D–G) on their General Certificate of Secondary Education (GCSE). GSCEs are a standardized examination taken at the end of compulsory education at 16 years of age. Of the analysis sample, 23% met criteria for poor educational qualifications.

NEET. NEET status was assessed at in-person interviews[58]. As of the age-18 interview, 12% of study members were NEET, similar to the UK population (as of 2010, about 14% of 19-year olds in the UK reported being NEET for at least 1 year[59].

Add Health. *Sample.* Add Health is an ongoing, nationally representative longitudinal study of the social, behavioural and biological linkages in health and developmental trajectories from early adolescence into adulthood. The cohort was drawn from a probability sample of 144 middle and high schools and is representative of American adolescents in grades 7–12 in 1994–1995. Since the start of the project, participants have been interviewed in-home at four data collection waves (numbered I–IV), most recently in 2007–2008, when 15,701 study members took part[25].

Genotyping. At the wave IV interview in 2007–2008, saliva and capillary whole blood were collected from respondents. Of the 15,701 individuals interviewed, 15,159 consented to genotyping and 12,254 agreed to genetic data archiving. DNA extraction and genotyping were conducted on this archive sample using two platforms (Illumina Omni1 and Omni2.5). After quality controls, genotype data were available for 9,975 individuals. We analysed data from $n = 5{,}690$ participants with genetically European ancestry. Imputation was conducted on SNPs called in more than 98% of the samples with a minor allele frequency of >1% using the Michigan Imputation Server (http://imputationserver.readthedocs.io/en/latest/pipeline/) and the Haplotype Reference Consortium reference panel[60].

Polygenic scoring. We computed polygenic scores for body mass index, schizophrenia and age at first birth following the method described by Dudbridge[20] according to the procedure used in previous studies[61]. Briefly, SNPs in the genotype database were matched to published GWAS results[16,17]. For each of these SNPs, a

loading was calculated as the number of phenotype-associated alleles multiplied by the effect size estimated in the original GWAS. Loadings were then averaged across the SNP set to calculate the polygenic score. The Add Health study was included in the most recent GWAS of educational attainment[18]. Thus, we obtained the polygenic score for educational attainment directly from the Social Science Genetic Association Consortium (SSGAC). The SSGAC computed the score according to the methods described in the GWAS article based on a GWAS that did not include any Add Health samples.

To account for any residual population stratification within the European-descent analysis sample, we residualized polygenic scores for the first 10 principal components estimated from the genome-wide SNP data[51] and standardized residuals to have mean = 0 and s.d. = 1 to compute polygenic scores for analysis. Principal components for the Add Health European-descent sample were provided by the SSGAC.

Neighbourhood characteristics. We measured neighbourhood-level socioeconomic disadvantage using Census-tract-level data linked to Add Health participants' addresses when they were first interviewed in 1994–1995 and when they were most recently followed-up in 2007–2008. Participants' addresses in 1994–1995 were linked with tract-level data from the 1990 Decennial Census[62]. Participants' addresses in 2007–2008 were linked with tract-level data from the 2005–2009 panels of the American Community Survey[63]. For each tract, we coded the proportions of female-headed households, individuals living below the poverty threshold, individuals receiving public assistance, adults with less than a high school education and adults who were unemployed using the following system: we computed tract deciles based on the full set of tracts from which Add Health participants were sampled at wave I. We then scored each tract on a scale of 1–10 corresponding to the wave I decile containing the tract's value on the variable. We calculated neighbourhood deprivation as the sum of decile scores across the 5 measures, resulting in a score ranging from 0 to 50. Values were Z-transformed to have mean = 0 and s.d. = 1 for analysis.

Add Health analysis included all European-descent Add Health participants with available genetic and neighbourhood data (n = 5,325).

Statistical analysis. We analysed continuous dependent variables using linear regression models. We analysed dichotomous dependent variables using Poisson regression models to estimate RRs. In models testing polygenic and neighbourhood risks for health and social problems, health and social problems were specified as the dependent variables and polygenic and neighbourhood risks were specified as predictor variables. We tested the statistical independence of polygenic risk information from family history risk information using multivariate regression with family history measures included as covariates alongside polygenic scores. In models testing for association between polygenic and neighbourhood risks, polygenic scores were specified as dependent variables and neighbourhood risks were specified as predictor variables. We tested whether associations between children's neighbourhood risks and polygenic risks were correlated because both were inherited from their parents using multivariate regression, with the mother's polygenic scores included as covariates alongside neighbourhood risk measures. We tested polygenic risk associations with neighbourhood mobility using the mobility model from previous work[64]; participants' adulthood neighbourhood disadvantage scores were regressed on their polygenic scores, their childhood neighbourhood disadvantage scores and covariates. For all models, we accounted for the non-independence of observations of siblings within families by clustering standard errors at the family level. For models testing polygenic score associations with neighbourhood conditions in the E-Risk data, only one member of each monozygotic twin pair was included in analysis. (For these models, monozygotic twins would have identical values for predictors and outcomes.) All models were adjusted for sex. Add Health models were adjusted for year of birth. (Year of birth did not vary in the E-Risk cohort.)

We conducted post-hoc power analyses to provide context for interpretation of the associations that we observed. We conducted power analysis using the 'power' command in the Stata software[65]. Both the E-Risk and Add Health samples had >80% power to detect associations with effect size r = 0.1 in all analyses. Power analysis for tests of polygenic score associations with neighbourhood risk is shown in Supplementary Fig. 3. We conducted power analysis for GCTA using the online power calculator provided by Hemani and Yang (http://cnsgenomics.com/shiny/gctaPower/). In the Add Health sample, power was >80% to detect SNP heritability estimates of ≥0.2. We did not conduct GCTA analysis in the E-Risk sample because power calculations suggested that only SNP heritabilities of >0.9 could be distinguished from 0 in that sample.

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

Data availability

The E-Risk data set reported in the current article is not publicly available owing to a lack of informed consent and ethical approval, but is available on request by qualified scientists. Requests require a concept paper describing the purpose of data access, ethical approval at the applicant's institution and provision for secure data access. We offer secure access on the Duke University and King's College London campuses. All data analysis scripts and results files are available for review. The Add Health data can be accessed through the Add Health study. Details are available through the Carolina Population Center as described here: https://www.cpc.unc.edu/projects/addhealth/documentation. Genotype data are available through dbGaP.

Code availability

All data analysis scripts and results files are available for review.

Received: 10 May 2018; Accepted: 19 February 2019;
Published online: 08 April 2019

References

1. Woods, L. M. et al. Geographical variation in life expectancy at birth in England and Wales is largely explained by deprivation. *J. Epidemiol. Community Health* **59**, 115–120 (2005).
2. Chetty, R. et al. The association between income and life expectancy in the United States, 2001–2014. *JAMA* **315**, 1750–1766 (2016).
3. Luo, Z.-C. et al. Disparities in birth outcomes by neighborhood income: temporal trends in rural and urban areas, British Columbia. *Epidemiology* **15**, 679–686 (2004).
4. Sampson, R. J. Urban sustainability in an age of enduring inequalities: advancing theory and ecometrics for the 21st-century city. *Proc. Natl Acad. Sci. USA* **114**, 8957–8962 (2017).
5. Newton, J. N. et al. Changes in health in England, with analysis by English regions and areas of deprivation, 1990–2013: a systematic analysis for the global burden of disease study 2013. *Lancet* **386**, 2257–2274 (2015).
6. Sampson, R. J., Morenoff, J. D. & Gannon-Rowley, T. Assessing 'neighborhood effects': social processes and new directions in research. *Annu. Rev. Sociol.* **28**, 443–478 (2002).
7. Oakes, J. M. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc. Sci. Med.* **58**, 1929–1952 (2004).
8. White, J. S. et al. Long-term effects of neighbourhood deprivation on diabetes risk: quasi-experimental evidence from a refugee dispersal policy in Sweden. *Lancet Diabetes Endocrinol.* **4**, 517–524 (2016).
9. Ludwig, J. et al. Long-term neighborhood effects on low-income families: evidence from moving to opportunity. *Am. Econ. Rev.* **103**, 226–231 (2013).
10. Chetty, R. & Hendren, N. *The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates* https://doi.org/10.3386/w23002 (National Bureau of Economic Research, 2016).
11. Chetty, R., Hendren, N. & Katz, L. F. The effects of exposure to better neighborhoods on children: new evidence from the moving to opportunity experiment. *Am. Econ. Rev.* **106**, 855–902 (2016).
12. Arcaya, M. C. et al. Role of health in predicting moves to poor neighborhoods among Hurricane Katrina survivors. *Proc. Natl Acad. Sci. USA* **111**, 16246–16253 (2014).
13. Oakes, J. M., Andrade, K. E., Biyoow, I. M. & Cowan, L. T. Twenty years of neighborhood effect research: an assessment. *Curr. Epidemiol. Rep.* **2**, 80–87 (2015).
14. Buka, S. L., Brennan, R. T., Rich-Edwards, J. W., Raudenbush, S. W. & Earls, F. Neighborhood support and the birth weight of urban infants. *Am. J. Epidemiol.* **157**, 1–8 (2003).
15. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
16. Schizophrenia Working Group of the Psychiatric Genomics Consortium Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **51**, 421–427 (2014).
17. Barban, N. et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
18. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
19. Odgers, C. L., Caspi, A., Bates, C. J., Sampson, R. J. & Moffitt, T. E. Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. *J. Child Psychol. Psychiatry* **53**, 1009–1017 (2012).
20. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
21. Belsky, D. W. et al. Polygenic risk, rapid childhood growth, and the development of obesity: evidence from a 4-decade longitudinal study. *Arch. Pediatr. Adolesc. Med.* **166**, 515–521 (2012).
22. Belsky, D. W. et al. Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: evidence from a 4-decade longitudinal study. *JAMA Psychiatry* **70**, 534–542 (2013).
23. Agerbo, E. et al. Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: a Danish population-based study and meta-analysis. *JAMA Psychiatry* **72**, 635–641 (2015).

24. Mujahid, M. S., Diez Roux, A. V., Morenoff, J. D. & Raghunathan, T. Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. *Am. J. Epidemiol.* **165**, 858–867 (2007).
25. Harris, K. M. et al. Social, behavioral, and genetic linkages from adolescence into adulthood. *Am. J. Public Health* **103**, S25–S32 (2013).
26. Yang, J. et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
27. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
28. Sariaslan, A. et al. Schizophrenia and subsequent neighborhood deprivation: revisiting the social drift hypothesis using population, twin and molecular genetic data. *Transl Psychiatry* **6**, e796 (2016).
29. Martin, J. et al. Association of genetic risk for schizophrenia with nonparticipation over time in a population-based cohort study. *Am. J. Epidemiol.* **183**, 1149–1158 (2016).
30. Gage, S. H., Smith, G. D. & Munafò, M. R. Schizophrenia and neighbourhood deprivation. *Transl Psychiatry* **6**, e979 (2016).
31. Sharkey, P. Neighborhoods, cities, and economic mobility. *RSF* **2**, 159–177 (2016).
32. Sharkey, P. *Stuck in Place: Urban Neighborhoods and the End of Progress Toward Racial Equality* (Univ. Chicago Press, 2013).
33. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
34. Polderman, T. J. C. et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
35. Hill, W. D. et al. Molecular genetic contributions to social deprivation and household income in UK Biobank. *Curr. Biol.* **26**, 3083–3089 (2016).
36. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
37. Martin, A. R. et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
38. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
39. Hartwig, F. P., Davies, N. M., Hemani, G. & Davey Smith, G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int. J. Epidemiol.* **45**, 1717–1726 (2016).
40. Trouton, A., Spinath, F. M. & Plomin, R. Twins Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Res. Hum. Genet.* **5**, 444–448 (2002).
41. Moffitt, T. E. & E-Risk Team Teen-aged mothers in contemporary Britain. *J. Child Psychol. Psychiatry* **43**, 727–742 (2002).
42. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
43. 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
44. Sherry, S. T. et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
45. Lapouse, R., Monk, M. A. & Terris, M. The drift hypothesis and socioeconomic differentials in schizophrenia. *Am. J. Public Health Nations Health* **46**, 978–986 (1956).
46. Murray, R. M., Jones, P. B., Susser, E., Os, J. V. & Cannon, M. *The Epidemiology of Schizophrenia* (Cambridge Univ. Press, 2002).
47. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).
48. Hamer, D. & Sirota, L. Beware the chopsticks gene. *Mol. Psychiatry* **5**, 11–13 (2000).
49. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.* **11**, 459–463 (2010).
50. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
51. Conley, D. et al. Assortative mating and differential fertility by phenotype and genotype across the 20th century. *Proc. Natl Acad. Sci. USA* **113**, 6647–6652 (2016).
52. Sampson, R. J., Raudenbush, S. W. & Earls, F. Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science* **277**, 918–924 (1997).
53. Sampson, R. J., Morenoff, J. D. & Earls, F. Beyond social capital: spatial dynamics of collective efficacy for children. *Am. Sociol. Rev.* **64**, 633–660 (1999).
54. *Waist Circumference and Waist–Hip Ratio: Report of a WHO Expert Consultation* http://www.who.int/nutrition/publications/obesity/WHO_report_waistcircumference_and_waisthip_ratio/en/ (WHO, 2008).
55. Scantlebury, R. & Moody, A. in *Health Survey for England, 2014* (eds. Craig, R., Fuller, E. & Mindell, J.) Vol. 1 Ch. 9 (The Health and Social Care Information Centre, 2014).
56. Schaefer, J. D. et al. Adolescent victimization and early-adult psychopathology: approaching causal inference using a longitudinal twin study to rule out noncausal explanations. *Clin. Psychol. Sci.* **6**, 352–371 (2018).
57. Caspi, A. & Moffitt, T. E. All for one and one for all: mental disorders in one dimension. *Am. J. Psychiatry* **175**, 831–844 (2018).
58. Goldman-Mellor, S. et al. Committed to work but vulnerable: self-perceptions and mental health in NEET 18-year olds from a contemporary British cohort. *J. Child Psychol. Psychiatry* **57**, 196–203 (2016).
59. Brown, J. *NEET: Young people Not in Education, Employment Or Training* (House of Commons, 2016).
60. The Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
61. Domingue, B. W. et al. The social genome of friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health. *Proc. Natl Acad. Sci. USA* **115**, 702–707 (2018).
62. Billy, J. O., Wenzlow, A. T. & Grady, W. R. *User Documentation for the Add Health Contextual Database* (Seattle Battelle Center Public Health Research, 1998).
63. Morales, L. & Monbureau, T. *Add Health Wave IV Contextual Data* (Add Health, 2013).
64. Belsky, D. W. et al. The genetics of success: how single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychol. Sci.* **27**, 957–972 (2016).
65. *Power and Sample Size* https://www.stata.com/features/power-and-sample-size/ (Stata, 2017).

Acknowledgements

The E-Risk Study is funded by the Medical Research Council (UKMRC grant G1002190). Additional support was provided by NICHD grant HD077482, Google and by the Jacobs Foundation. The Add Health study is supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development grants P01HD31921, R01HD073342 and R01HD060726, with cooperative funding from 23 other federal agencies and foundations. D.W.B. and C.L.O. were supported by fellowships from the Jacobs Foundation. C.L.O. is supported by the Canadian Institute for Advanced Research. B.W.D. is supported by the Russell Sage Foundation award 961704. We are grateful to the E-Risk study mothers and fathers, the twins and the twins' teachers, and the Add Health study participants and their parents for their participation. Our thanks to CACI, Google Street View and to members of the E-Risk team for their dedication, hard work and insights. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

D.W.B., A.C., T.E.M. and C.L.O. designed the research. A.C., T.E.M., L.A., C.L.O. and K.M.H. collected the data. Data were analysed by D.W.B., B.W.D., R.M.H., D.L.C. and J.P. All authors reviewed drafts and provided critical feedback and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0562-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.W.B. or C.L.O.

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

nature research

Corresponding author(s): Daniel W. Belsky
Cadice Odgers

Last updated by author(s): 1-24-2019

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Data collection was from two prospective longitudinal cohort studies and comprised in-person interviews to obtain phenotypic information and DNA collection, extraction, and analysis. Documentation for DNA handling is included in the text or referenced appropriately.
Data analysis	Data were analyzed using the Stata an R softwares.

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The E-Risk dataset reported in the current article is not publicly available due to lack of informed consent and ethical approval, but is available on request by qualified scientists Requests require a concept paper describing the purpose of data access, ethical approval at the applicant's institution, and provision for secure data access. We offer secure access on the Duke University, and King's College London campuses. All data analysis scripts and results files are available for review. The Add Health data can be accessed through the Add Health Study. Details are available through the Carolina Population Center as described here: https://www.cpc.unc.edu/projects/addhealth/documentation. Genotype data are available through dbGaP.

Field-specific reporting

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

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

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

Behavioural & social sciences study design

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

Study description	Quantitative analysis of data collected in prospective longitudinal studies.
Research sample	Data were drawn from two studies. The Environmental Risk (E-Risk) Longitudinal Twin Study tracks the development of a birth cohort of 2,232 British children. The sample was drawn from a larger birth register of twins born in England and Wales in 1994–1995 41. Full details about the sample are reported elsewhere (manuscript ref 42). Briefly, the E-Risk sample was constructed in 1999–2000, when 1,116 families (93% of those eligible) with same-sex 5-year-old twins participated in home-visit assessments. The sample includes 56% monozygotic and 44% dizygotic twin pairs; sex is evenly distributed within zygosity (49% male). Families were recruited to represent the UK population of families with newborns in the 1990s, on the basis of residential location throughout England and Wales, and mother's age. Teenaged mothers with twins were over-selected to replace high-risk families who were selectively lost to the register through nonresponse. Older mothers having twins via assisted reproduction were under-selected to avoid an excess of well-educated older mothers. These strategies ensured that the study sample represents the full range of socioeconomic conditions in Great Britain (manuscript ref 19). Follow-up home visits were conducted when the children were aged 7 (98% participation), 10 (96% participation), 12 (96% participation), and, in 2012–2014, 18 years (93% participation). There were no differences between those who did and did not take part at age 18 in terms of socioeconomic status (SES) assessed when the cohort was initially defined ($\chi 2 = 0.86$, $p = .65$), age- 5 IQ scores ($t = 0.98$, $p = .33$), or age- 5 internalizing or externalizing behavior problems ($t = 0.40$, $p = .69$ and $t = 0.41$, $p = .68$, respectively). Home visits at ages 5, 7, 10, and 12 years included assessments with participants as well as their mother; the home visit at age 18 included interviews only with the twin participants. All interviews at the age-18 assessment were conducted after the 18th birthday. Each twin participant was assessed by a different interviewer. The joint Research and Development Office of South London and Maudsley and the Institute of Psychiatry Research Ethics Committee approved each phase of the study. Parents gave informed consent and twins gave assent between ages 5 and 12 years; twins gave informed consent at age 18 years. The National Longitudinal Study of Adolescent to Adult Health (Add Health) is an ongoing, nationally-representative longitudinal study of the social, behavioral, and biological linkages in health and developmental trajectories from early adolescence into adulthood. The cohort was drawn from a probability sample of 144 middle and high schools and is representative of American adolescents in grades 7-12 in 1994-1995. Since the start of the project, participants have been interviewed in home at four data collection waves (numbered I-IV), most recently in 2007-2008, when 15,701 Study members took part (see ref 25 of the article for details).
Sampling strategy	We analyzed data from all E-Risk and Add Health participants for whom genetic and neighborhood data were available.
Data collection	E-Risk data were collected through in-person interviews, including anthropometric assessments, and blood draws for DNA collection. Add Health data were collected through in-person interviews and saliva DNA collection.
Timing	E-Risk data analyzed in this article were collected during 1999-2014. Add Health data analyzed in this article were collected during 1994-2008.
Data exclusions	We analyzed data from European-descent participants with available genetic and neighborhood data.
Non-participation	NA
Randomization	NA

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above.
Recruitment	Paper describes secondary data analysis
Ethics oversight	Ethics oversight for the research reported in this article was provided by Institutional Review Boards at Duke University and the University of North Carolina at Chapel Hill.

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