Effects of Genetic Propensity for Education on Labor Market and Health Trajectories across the Working Life*

Stefano Lombardi† [1,2,3,4], Nurfatima Jandarova† [5], Kristina Zguro† [6], Jarkko Harju[5], Aldo Rustichini[7], and Andrea Ganna[6,8,9]

[1]VATT Institute for Economic Research, Helsinki, Finland
[2]IFAU and Uppsala Center for Labor Studies, Uppsala, Sweden
[3]IZA, Bonn, Germany
[4]Rockwool Foundation, Berlin, Germany
[5]Tampere University and Finnish Centre of Excellence in Tax Systems Research (FIT), Tampere, Finland
[6]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Finland
[7]University of Minnesota, MN, USA
[8]Broad Institute of MIT and Harvard, MA, USA
[9]Analytic and Translational Genetics Unit, Massachusetts General Hospital, MA, USA

Abstract

Education is a major source of inequality in income and health. Polygenic indices for educational attainment (EA-PGI) capture both direct and indirect genetic influences on education, but their effects on income and health trajectories remain unclear. Using Finnish registry data on 51,056 graduates followed annually since graduation for up to 25 years, we report three findings. First, higher EA-PGI strongly predicts income growth, but only among higher-educated people: tertiary-educated graduates at the 90th percentile earn €45,392 (13.1 %) higher discounted lifetime income than those at the 10th percentile. This effect is not mediated by overall health. Second, EA-PGI does not predict income differences at labor market entry or the quality of the first employer, but rather a higher job-to-job mobility toward better-paying firms, which drives the long-run income divergence. Third, controlling for parental EA-PGI in 12,871 parent–offspring trios reduces the discounted lifetime income gap by 71 %, and the effect of paternal (but not maternal) EA-PGI on offspring income exceeds that of the offspring's own EA-PGI. These findings suggest that genetic factors associated with educational attainment predict income trajectories primarily through faster and more frequent changes to higher-paying employers. However, much of this association reflects indirect paternal genetic effects, consistent with enduring paternal patterns of intergenerational job and income transmission.

*Corresponding authors: Andrea Ganna (andrea.ganna@helsinki.fi), Stefano Lombardi (stefano.lombardi@vatt.fi).

†Co-first authors.

Introduction

Understanding the origins and persistence of differences in health, education, and income – collectively referred to as *socioeconomic status* – is of central importance to policymakers and society. While individual effort and choice contribute to these disparities, they can also be largely shaped by life circumstances. For a child being born into an affluent family or having a low genetic predisposition to disease is a matter of chance. However, such factors shape socioeconomic status and, consequently, can affect equality of opportunity in society (Roemer and Trannoy, 2016).

Economists and epidemiologists have long sought to identify the drivers of socioeconomic inequality (for health disparities, see e.g. Chetty et al., 2016; Schwandt and Von Wachter, 2020; Dwyer-Lindgren et al., 2022; Bundy et al., 2023). Income and health outcomes arise from a complex interplay between genetics and environmental factors over the life cycle. This idea is central to economic models of skill formation (Cunha and Heckman, 2007), in which genetic endowments are propensities that occur via environmental interactions (Biroli et al., 2025). Only recent developments in molecular genetics, however, have made it possible to directly study how specific genetic endowments relate to socioeconomic status.

The availability of polygenic indices (PGI) – summary measures of genetic propensities for complex traits – has opened new avenues for studying the role of genetics in socioeconomic outcomes. Most human traits exhibit substantial heritability (Barth, Papageorge, and Thom, 2020; Harden and Koellinger, 2020; Rustichini et al., 2023; Carvalho, 2025), with educational attainment (EA) being a primary focus of this growing literature thanks to large-scale genome-wide association studies (GWAS) and validated PGIs (Okbay et al., 2022; Abdellaoui et al., 2023). These scores capture both direct genetic effects – randomly inherited at conception – and indirect effects, such as environmentally mediated influences of parental genetics (Kong et al., 2018; Sohail et al., 2019; Sjaarda and Kutalik, 2023; Tan et al., 2024). Understanding how environmental factors mediate genetic predispositions is key to identifying when interventions can amplify or offset these influences, and this remains an active area of research.

This project offers a novel perspective on how genetic factors contribute to socioeconomic inequalities. To our knowledge, this is the first large study to combine EA-PGIs with long-term registers to investigate how genetics influences income trajectories through sorting into distinct educational pathways and into firms with differing productivity. Although a large literature in economics emphasizes the role of employers in shaping wage inequality (Kline, 2024), the role of firms in mediating genetic effects on income remains unknown. In addition, most studies that use PGIs rely on cross-sectional data, providing static estimates that are silent about the dynamics of inequality. One recent exception is Akimova et al. (2025), who analyze career trajectories in the UK biobank. It also remains unclear whether genetic effects are driven by certain population subgroups, such as by low- or high-educated individuals.

To fill this gap, our analysis combines comprehensive Finnish administrative register data on education, income, demographics, and health with EA-PGI. Our primary labor market outcome, labor income, is drawn from continuous tax records, avoiding self-report bias, missing-not-at-random mechanisms, and other measurement errors that could otherwise cloud the results. Each fresh graduate is repeatedly observed throughout their prime working ages since graduation until up to 25 years later, and workers are matched each year to their current employer.

This matched employer–employee panel data allows us to implement a trajectory-based approach to examine how the genetic predisposition to education influences the evolution of labor income over the life course, both on average and across educational groups. We additionally leverage detailed measures of employer and employee productivity to characterize the labor-market dynamics through which individuals with different EA-PGIs achieve divergent income trajectories. We further examine how labor income is shaped by overall health and conduct within-family analyses that control for parental PGIs to disentangle direct from indirect genetic effects.

Results

EA-PGI predicts lifetime income trajectories, but only among individuals with tertiary education

The study includes 51,056 individuals with genome-wide genetic data linked to longitudinal health and socioeconomic information. The data covers employment histories (i.e., employee-employer links with job spells length), annual income from labor, and educational records (education level, field and school/university identifiers). This information spans thirty years (1987–2019) and all individuals are followed from graduation up to 25 years later. We estimate a dynamic model to analyze the relation between EA-PGI and individual income trajectories over time, adjusting for calendar year, birth year, gender, the first ten genetic principal components, and biobank indicator.

Individuals included in this study are generally better educated and with a higher share of women compared to the nationwide population of fresh graduates in the same calendar period (Supplementary Table A.2a). Reassuringly, when we reweight the sample to reflect the characteristics in the population the results are virtually unaffected (Supplementary Table A.11).

We first confirm that EA-PGI is significantly associated with years of educational attainment, with results comparable to what previously reported in the literature (R2=7.1 %, Supplementary Table A.12) (Wang et al., 2021). Comparing earned income trajectories between the 10th and 90th percentile of the EA-PGI distribution over 25 years since graduation, we show that income levels are initially very similar across groups (Figure 1a). However, trajectories begin to diverge substantially over time. The gap in average annual income between 90th and 10th percentile of EA-PGI 10 years from graduation is €3,262 [CI €2,741 - €3,783] and widens further to €7,534 [CI €6,382 - €8,687] by 25 years from graduation. To put this in context, the average yearly earnings in the sample is €24,932 and €34,858 at 10 and 25 years after graduation, respectively. The cumulated income over the 25 years after graduation, discounted to obtain its present value at graduation, is €309,659 [CI €307,099 - €312,219] for individuals at the 10th percentile and €350,418 [CI €347,300 - €353,536] for those at the 90th percentile of the EA-PGI distribution (Table 1). The gap in cumulated income between the 10th and 90th percentile is €40,759, equivalent to 13.1 % and corresponding to 14 % of the median income over the same period.

Strikingly, this genetic gradient in income varies markedly by educational attainment. The income gap by genetics is driven entirely by individuals with tertiary education (Figures 1b and 1c), while no systematic differences are observed among the people with secondary education (63.4 % of the graduates has tertiary degree). Among the tertiary-educated, differences in income trajectories widen up during the first 15 years after graduation and then stabilize, at which point individuals are, on average, 40 years old and approaching peak labor market attachment.

Such a differential result likely reflects two forces: individuals with relatively high EA-PGI experience higher economic returns to ability, and EA-PGI influences educational sorting, shaping the types and levels of education they pursue. Although we cannot fully separate sorting from the economic returns to higher EA-PGI, we next analyze how EA-PGI correlates with a measure of ability and examine the role of employers in explaining the observed income differences.

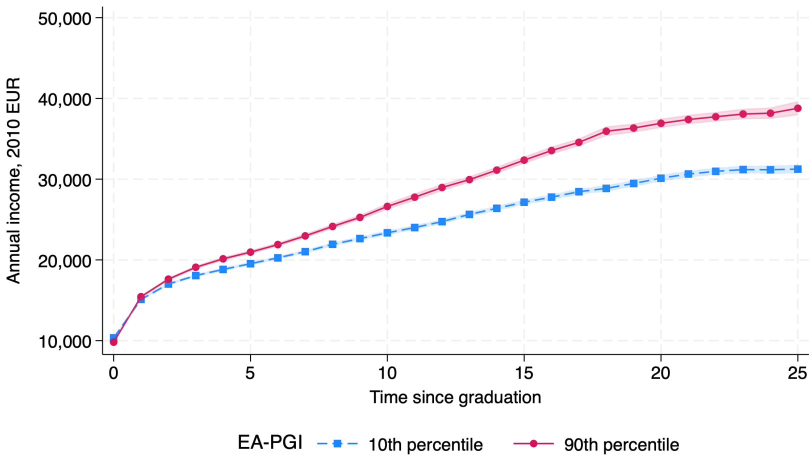
EA-PGI is associated with the tertiary-educated workers' productivity

To understand the mechanisms through which the EA-PGI influences income trajectories, we first examine its relationship with both employee and employer productivity in the labor market. To quantify productivity, we use full population data within an Abowd–Kramarz–Margolis (AKM) framework, the workhorse model in economics for decomposing wage variation into worker and firm components (Abowd, Kramarz, and Margolis, 1999; Kline, 2024).

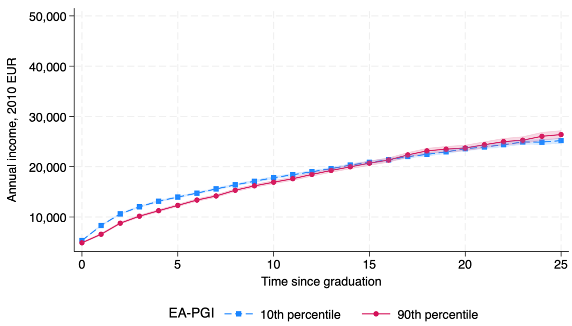
The model exploits repeated measurements of worker's wage, relating it to individual- and firm-specific indicators. After model estimation, each worker is assigned a corresponding estimated intercept, which we refer to as *worker productivity index*, as it cap-

Figure 1: Average annual income by EA-PGI level, over time and by education
Panel (a) uses full analysis sample ($N = 51{,}056$), while Panels (b) and (c) use subset of workers based on their highest qualification being either secondary ($N = 18{,}692$) or tertiary degree ($N = 32{,}364$), respectively. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Average income estimated from a regression of annual income on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95 % CIs.

(a) Pooled



(b) Secondary



(c) Tertiary

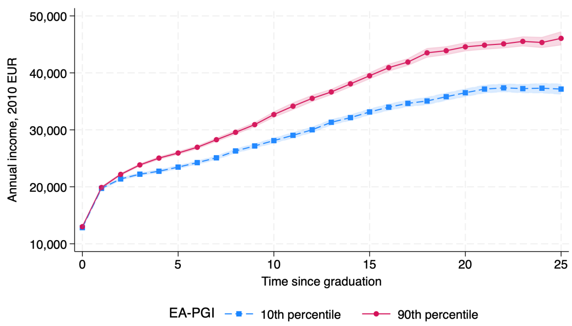


Table 1: Cumulated lifetime income by EA-PGI level
The table reports adjusted lifetime income (up to 25 years since graduation) by EA-PGI percentiles. Column (1) uses full analysis sample ($N = 51{,}056$), while columns (2) and (3) use subset of workers based on their highest qualification being either secondary ($N = 18{,}692$) or tertiary degree ($N = 32{,}364$), respectively. Average lifetime income adjusted by regressing cumulated income on EA-PGI and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Standard errors reported in parentheses.

	Dependent variable: Cumulated income		
	Pooled (1)	Secondary (2)	Tertiary (3)
EA-PGI percentiles			
10th	309 659	262 386	346 194
	(1 306)	(1 429)	(1 944)
20th	316 476	260 040	353 786
	(1 037)	(1 153)	(1 539)
30th	321 505	258 309	359 386
	(898)	(1 046)	(1 304)
40th	325 804	256 829	364 174
	(844)	(1 042)	(1 175)
50th	329 893	255 422	368 728
	(857)	(1 116)	(1 137)
60th	333 908	254 041	373 198
	(930)	(1 248)	(1 188)
70th	338 222	252 556	378 004
	(1 061)	(1 439)	(1 328)
80th	343 389	250 777	383 758
	(1 265)	(1 709)	(1 579)
90th	350 418	248 358	391 585
	(1 591)	(2 120)	(2 006)
Obs.	51 056	18 692	32 364

tures the persistent component of the worker's wage that is portable when switching jobs across firms, net of the effect of firm-quality and time-varying worker's characteristics.

We find a statistically significant correlation between the worker productivity index estimated via the AKM model and EA-PGI (both standardized to have mean 0 and standard deviation 1), but primarily for the tertiary-educated workers (Figure 2). For this group, a one standard deviation increase in EA-PGI is associated with a 0.115 [CI 0.101 - 0.129] standard deviation higher worker productivity, compared to a 0.017 [CI 0.003 - 0.031] increase among the secondary-educated individuals. Hence, individuals with high EA-PGI tend to be highly productive on the labor market (and therefore are on average paid higher wages), as long as they obtain a tertiary education degree.

We further examine whether the correlation between EA-PGI and worker productivity persists across education groups after adjusting for fine-grained education-related characteristics (education field, and academic institution). Although the association between EA-PGI and worker productivity attenuates after adjustment (Supplementary Table A.6), it remains statistically significant among tertiary-educated workers and is significantly stronger than among those with secondary education.

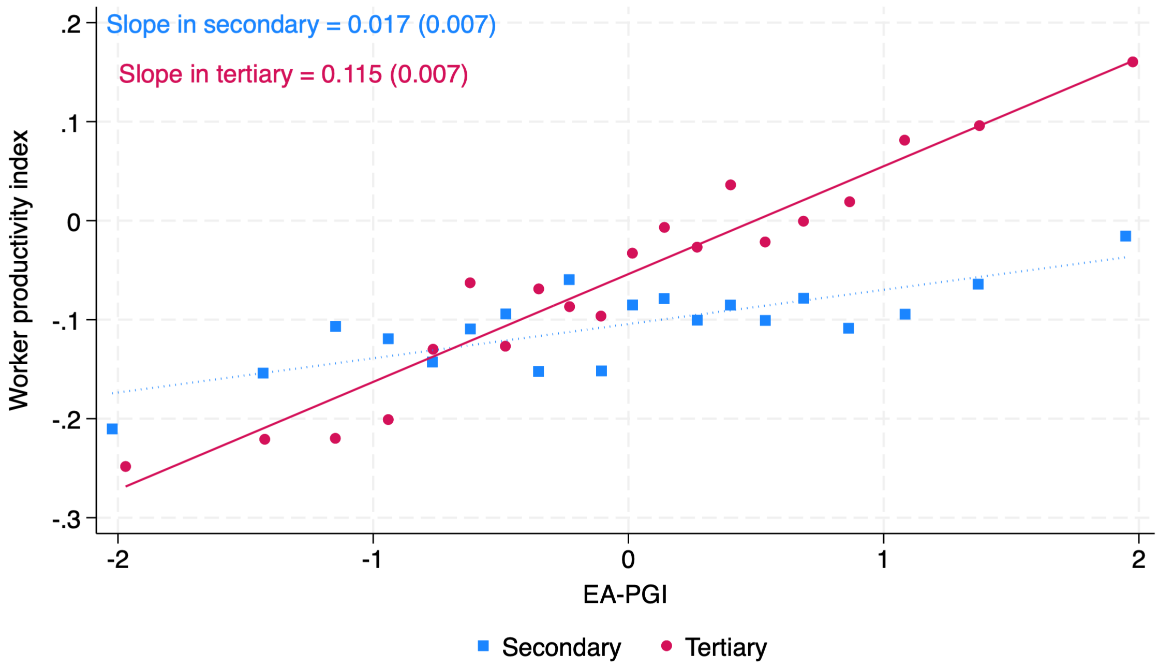
High EA-PGI individuals transition more rapidly and more frequently to higher-quality firms

We next investigate the role of firms in shaping income differences across levels of EA-PGI among tertiary-educated individuals. Workers with higher EA-PGI change employers slightly more frequently (4.3 jobs vs 4.5 jobs over 25 years, P-value=0.007 for the 10th and 90th percentile EA-PGI, respectively; Figure 3a). To assess the quality dimension of these moves, we use the AKM regression to estimate a *firm quality index* which captures the firm-specific component of paid wage that is common to all workers at that firm, after accounting for worker ability and observed characteristics. We use the firm quality index to rank all firms in which workers in the sample are employed, and use this as an outcome in our trajectories model.

We show that, on average, higher EA-PGI individuals transition to significantly higher-quality firms (Figure 3b). Interestingly, the quality of their first employer is not statistically different across the EA-PGI distribution, but the firm quality gap widens as early as three years after graduation. This pattern suggests that individuals do not initially self-select into higher- or lower-quality firms based on their EA-PGI, but sorting by EA-PGI appears to begin early in the career, with higher EA-PGI individuals progressively moving toward better-quality firms.

In contrast, among individuals with secondary education, firm quality trajectories appear similar across EA-PGI levels (Supplementary Figure A.3). These findings sug-

Figure 2: Worker productivity index and EA-PGI level by education
The vertical axis reports a worker productivity index estimated via an AKM regression (details reported in Section 1.2); the horizontal axis reports ventiles of the EA-PGI distribution. The binscatter plot shows the relation between these two quantities standardized to have mean 0 and standard deviation 1 and after having residualized them with respect to gender, year of birth indicators, and first 10 genetic principal components (PCs). The data used to obtain the plot is a cross-section of 31,866 individuals in the analysis sample with non-missing worker productivity index. The lines correspond to sub-samples based on highest education achieved. The figure also reports the estimated slopes (and standard errors in parentheses) from the corresponding linear regression of worker productivity on EA-PGI controlling for gender, year of birth, calendar year and biobank indicators, and first ten genetic PCs.

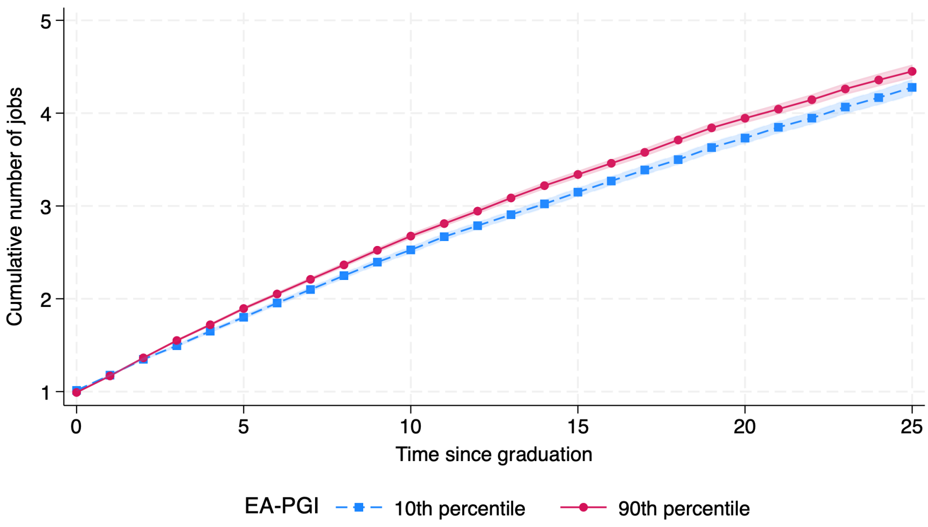


gest that the higher income trajectories observed among tertiary-educated individuals with high EA-PGI are largely driven by greater access to higher-quality, higher-paying firms over time, rather than by difference in educational attainment per se, initial labor market entry, or the frequency of employment transitions alone.

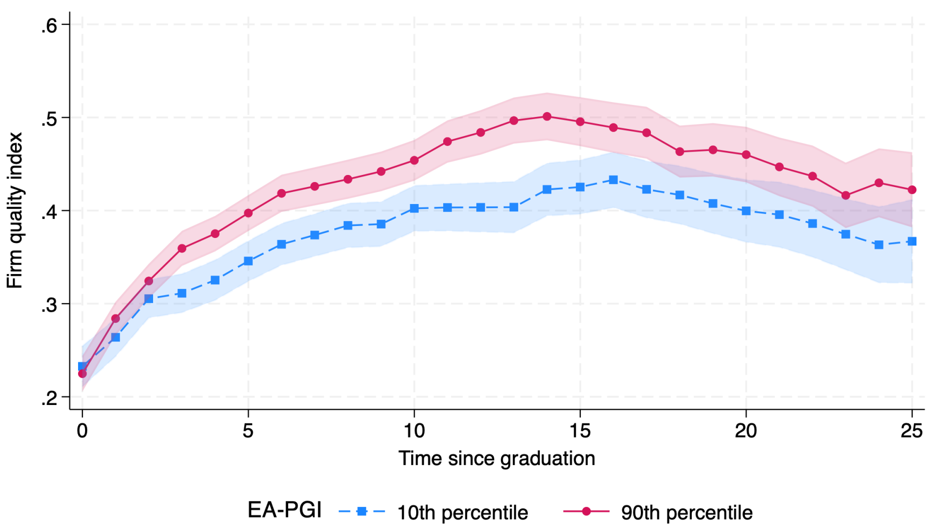
Figure 3: Employer mobility and quality of tertiary-educated individuals by EA-PGI level, over time

Panel **(a)** plots the average number of employment spells, and Panel **(b)** - the average firm quality index over time since graduation. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Both panels are estimated from a regression of respective outcome on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The estimation sample is restricted to tertiary-educated workers with non-missing firm identifiers ($N = 31{,}015$) in panel **(a)** and non-missing firm quality index ($N = 22{,}733$) in panel **(b)**. The shaded areas correspond to 95 % CIs.

(a) Cumulative number of jobs since graduation



(b) Firm quality since graduation



Income growth disparities between high and low EA-PGI are attributable to differences in mobility between rather than within jobs

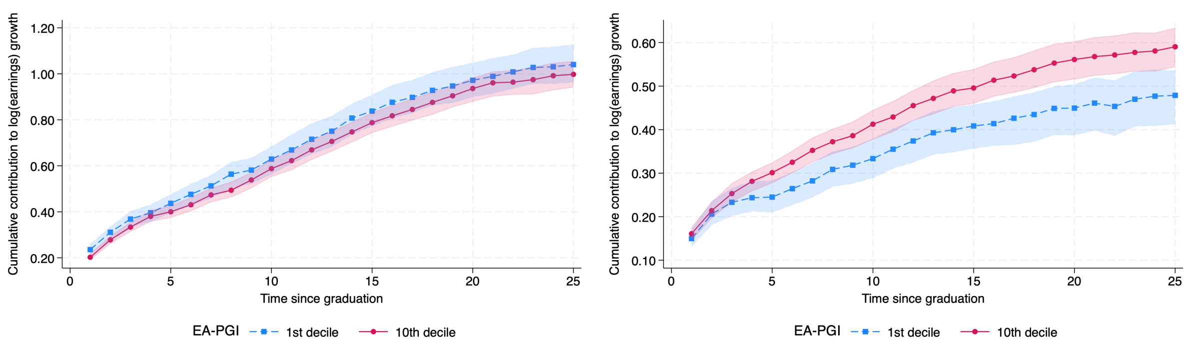
Having established that tertiary-educated individuals with higher EA-PGI tend to transition more rapidly to higher-quality firms, we next examine how much of the income differences between EA-PGI groups can be attributed to mobility within versus between firms. To do so, we decompose changes in labor income between consecutive years since graduation into within-firm wage growth (stayers), employer-to-employer mobility, and transitions into and out of non-employment (Hahn, Hyatt, and Janicki, 2021).

Overall, within-firm wage growth contributes most to cumulative income gains over time and does so similarly for individuals at the 10th and 90th percentiles of EA-PGI (Figure 4a). However, the contribution of between-firm mobility to earnings growth becomes increasingly important over time for those at the 90th percentile compared to those at the 10th percentile (Figure 4b). After 25 years, the cumulative job-to-job contribution to earnings growth differs by approximately 0.1 log-points (about 10 %) between the two groups, suggesting that a substantial portion of the income divergence is driven by firm-to-firm mobility. Transitions into or out of non-employment account for only a negligible share of these disparities (Supplementary Figure A.4).

Figure 4: Decomposition of cumulative growth in log annual income of tertiary-educated individuals by EA-PGI level, over time

Panel **(a)** reports cumulative contribution of within-firm earnings growth and Panel **(b)** - of employer-to-employer mobility to overall cumulative log earnings growth over time since graduation. The blue line corresponds to 1st and the red - to 10th decile of EA-PGI distribution. The decomposition follows Hahn, Hyatt, and Janicki (2021). The sample for the decomposition is restricted to tertiary-educated workers ($N = 32{,}364$). The shaded areas correspond to 95 % CIs obtained via 500 block bootstrap iterations, where at each iteration we sample with re-immission 32,364 whole income histories from the pool of tertiary graduates.

(a) Stayers **(b) Employer-to-employer movers**

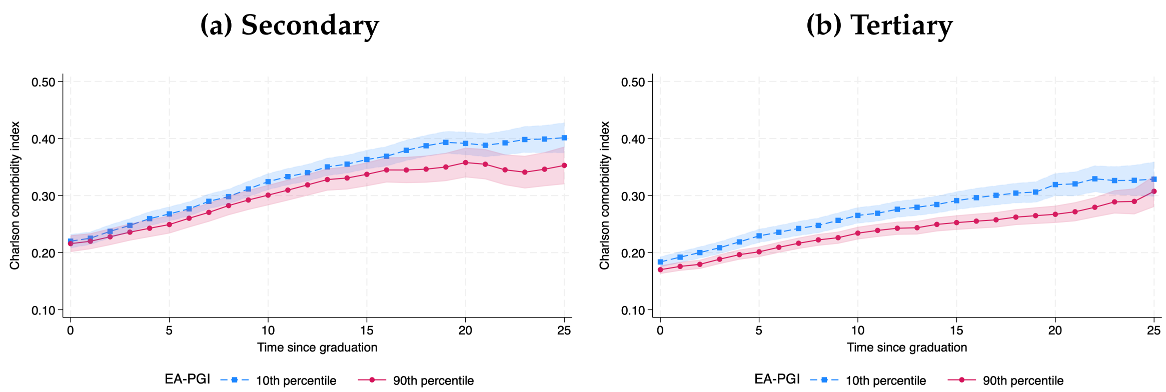


Overall health contributes similarly to the EA-PGI–income association for secondary- and tertiary-educated individuals

Health is a well-established determinant of educational (Pallesen et al., 2024; Newman, Gordon, and Mendes, 2025) and income disparities, making it a potential intermediate channel between EA-PGI and income. To examine whether the correlation between EA-PGI and income is mediated by disease burden, we calculate the Charlson Comorbidity Index (CCI), which captures the first occurrence of 17 major chronic conditions over the life course (Charlson et al., 1987; Deyo, Cherkin, and Ciol, 1992) and use it as an outcome in our trajectories regression.

We find that the CCI is, on average, lower among individuals with tertiary compared to secondary education (Figure 5), reflecting a well-established lower incidence of major chronic diseases among high-educated individuals (Agardh et al., 2011; Tillmann et al., 2017; Vaccarella et al., 2023). Consistent with this pattern, higher EA-PGI is significantly associated with a lower cumulative disease burden as measured by the CCI. The magnitude of this association is highly similar across education groups, suggesting that the stronger association between EA-PGI and income observed among tertiary-educated individuals is not primarily explained by differences in disease burden. This result is confirmed when controlling for parental EA-PGI (Supplementary Figure A.5).

Figure 5: Average health index by EA-PGI level, over time and by education
Panels **(a)** and **(b)** report the results for secondary- (N = 18,669) and tertiary-educated workers (N = 32,294) with non-missing Charlson Comorbidity Index, respectively. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Average health index estimated from a regression of Charlson Comorbidity Index on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95 % CIs.



Parental EA-PGI, in particular that of fathers, predicts the income trajectories among tertiary-educated people

EA-PGI captures both direct genetic effects and indirect effects, such as environmentally mediated influences from parental genetics (i.e., genetic nurture), as well as population stratification and assortative mating (Kong et al., 2018). Indeed, the people at the bottom decile of EA-PGI show remarkably lower socioeconomic status (parental education) than those at the top decile (Supplementary Table A.5). To better isolate direct genetic effects, we leverage parental genetic data. Specifically, we calculate maternal and paternal EA-PGI using SNIPAR (Young, Nehzati, Benonisdottir, et al., 2022) for 12,871 parent–offspring trios. Of these, 4,586 were directly genotyped, while the remainder were imputed based on 4,482 duos and 3,803 sibling pairs.

We first confirm that the association between offspring EA-PGI and years of education is attenuated by approximately 25 % after accounting for both maternal and paternal EA-PGIs (see e.g. Young, Nehzati, Benonisdottir, et al., 2022), and that both parental EA-PGIs are significantly associated with offspring education (Supplementary Table A.7), with effects of similar magnitude.

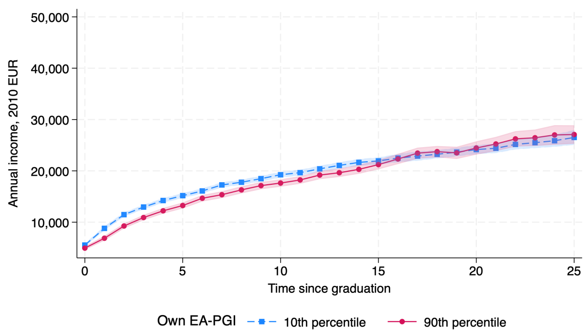
Next, we estimate the impact of offspring EA-PGI on income trajectories while adjusting for both maternal and paternal EA-PGIs. Controlling for parental EA-PGI accounts for income variation attributable to parental genetics as well as family background characteristics unevenly distributed across EA-PGI groups (Rustichini et al., 2023). Among secondary-educated individuals, the income gap across EA-PGI percentiles remains negligible, consistent with our earlier findings (Figure 6a). In contrast, among tertiary-educated individuals, adjusting for parental EA-PGI substantially reduces the income gap between the 90th and 10th percentiles of offspring EA-PGI by approximately 71 % (Figure 6b vs 6d; Table 2; Supplementary Tables A.8 and A.10).

However, unlike what was observed for years of education, only paternal (and not maternal) EA-PGI is significantly associated with offspring income trajectories (Supplementary Table A.8 and Supplementary Figure A.6). Strikingly, the effect of paternal EA-PGI on offspring income is larger than that of the offspring's own EA-PGI (€12,113 [CI €5,509 - €18,717] vs €5,048 [CI €−2,466 - €12,561] per one standard deviation in respective EA-PGI). To contextualize, regardless of offspring EA-PGI, fathers in the 10th percentile of EA-PGI are associated with a cumulative offspring income 25 years after graduation of €330,699 [CI €321,958 - €339,440], compared with €360,946 [CI €350,897 - €370,994] among fathers in the 90th percentile. These effects are consistent across both directly genotyped and imputed trios (Supplementary Table A.9).

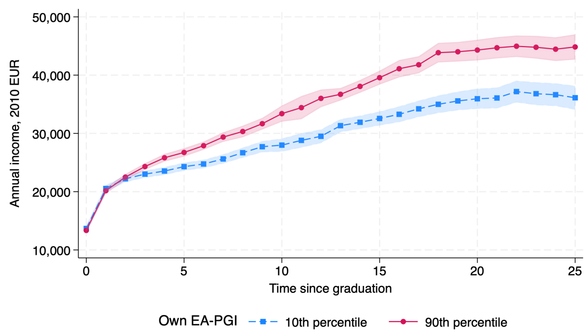
Figure 6: Average annual income by EA-PGI level, over time and by education, unconditional and conditional on parental EA-PGI among parent-offspring trios

Panels **(a)** and **(b)** plot baseline trajectories among secondary- and tertiary-educated workers, respectively, without controlling for parental EA-PGI. Panels **(c)** and **(d)** plot the trajectories among secondary- and tertiary-educated workers after controlling for parental EA-PGI fully interacted with time since graduation. The estimation sample in Panels **(a)** and **(c)** is restricted to secondary- ($N = 5{,}063$) and in Panels **(b)** and **(d)** - to tertiary-educated ($N = 7{,}808$) workers with non-missing parental EA-PGI. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Average income estimated from a regression of annual income on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95 % CIs.

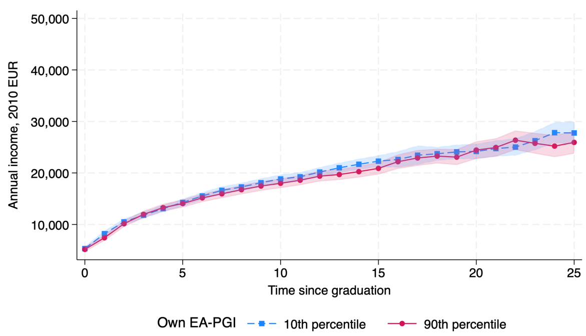
(a) Baseline without parental PGI (secondary)



(b) Baseline without parental PGI (tertiary)



(c) Controlling for parental PGI (secondary)



(d) Controlling for parental PGI (tertiary)

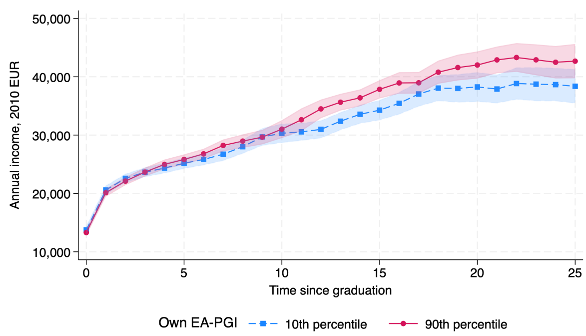


Table 2: Cumulated lifetime income of tertiary-educated individuals by EA-PGI level, conditional on parental EA-PGI among parent-offspring trios
The table reports adjusted lifetime income (up to 25 years since graduation) by EA-PGI percentiles. The estimation sample is restricted to parent-offspring trios with tertiary-educated offspring ($N = 7{,}808$). Column (1) reports the estimates in the baseline specification without controlling for parental EA-PGI. Columns (2)-(4) report average cumulated lifetime income by own, maternal and paternal EA-PGI percentiles, respectively, conditional on parental EA-PGI. Average lifetime income adjusted by regressing cumulated income on EA-PGI and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Standard errors reported in parentheses.

	Dependent variable: Cumulated income			
	Baseline	Controlling for parental EA-PGI		
	Own EA-PGI (1)	Own EA-PGI (2)	Mother EA-PGI (3)	Father EA-PGI (4)
EA-PGI percentiles				
10th	325 796	339 181	339 879	330 699
	(4 015)	(5 690)	(4 311)	(4 459)
20th	332 743	341 428	341 893	335 844
	(3 160)	(4 186)	(3 286)	(3 313)
30th	337 737	343 044	343 306	339 552
	(2 677)	(3 232)	(2 713)	(2 664)
40th	341 952	344 408	344 571	342 722
	(2 413)	(2 611)	(2 388)	(2 338)
50th	345 756	345 639	345 677	345 561
	(2 328)	(2 320)	(2 313)	(2 309)
60th	349 747	346 930	346 834	348 447
	(2 413)	(2 399)	(2 466)	(2 543)
70th	354 053	348 323	348 007	351 537
	(2 685)	(2 884)	(2 823)	(3 019)
80th	359 280	350 014	349 535	355 397
	(3 197)	(3 799)	(3 487)	(3 808)
90th	365 948	352 172	351 466	360 946
	(4 022)	(5 198)	(4 500)	(5 126)
Obs.	7 808	7 808	7 808	7 808

Discussion

In this study, we describe the impact of known genetic factors associated with educational attainment on socio-economic inequalities, and particularly labor income. To our knowledge, this is the first large study to investigate the impact of EA-PGI on longitudinal income trajectories; prior work has been limited to small samples or cross-sectional designs. It is important to recognize that these studies have contributed to an aversion, anger and even fear of studying genetics in social stratification research (Martschenko, Trejo, and Domingue, 2019; Mills, 2022). We include a document with Frequently Asked Questions (FAQs) to offer an accessible explanation of what our study does and, importantly, does not, find and how the results should be interpreted.

We focus on labor income because it represents a well-defined and downstream indicator of socioeconomic status, capturing differences in education, health, and labor-market skills. Labor income, together with government transfers, constitutes the main source of monetary inflows accumulated over the life course, making it a fundamental component of wealth (Black et al., 2025). Moreover, income inequality is a central topic in both economics (Kline, 2024) and epidemiology (Lynch et al., 2000; Pickett and Wilkinson, 2015; Schwandt et al., 2021), allowing us to build on well-established methodologies while incorporating genetics into the same analytical framework. Finally, income is strongly correlated with health, which is the other primary component of socioeconomic status.

We show that workers earn similar income upon labor market entry, irrespective of their EA-PGI level. After entry, however—and only among tertiary-educated individuals—income differences between those at the 90th and 10th percentiles of EA-PGI widen steadily over time, culminating in a 14 % cumulative income gap 25 years after graduation. The fact that EA-PGI predicts income trajectories only among tertiary-educated individuals likely reflects two forces. First, individuals with relatively high EA-PGI experience higher economic returns to ability. Second, EA-PGI influences educational sorting itself, shaping the types and levels of education pursued. While we cannot quantify exactly the relative importance of the two channels (sorting and economic returns to higher EA-PGI), our analysis based on the AKM model estimation (Abowd, Kramarz, and Margolis, 1999) offers additional insights on the role of employers in explaining the observed income differences.

We show that a measure of labor market productivity (or ability) is indeed positively associated with EA-PGI, but again only for tertiary-educated workers. This relationship persists even after controlling for detailed education-related characteristics (education field and type of academic institution). The finding that EA-PGI relates to labor market productivity only among highly educated individuals raises the question of whether some high-ability individuals with only secondary education might benefit from additional educational support or incentives to pursue tertiary studies. While our data cannot directly address this, we note that any such discussion must consider the role of family socioeconomic background (Ichino, Rustichini, and Zanella, 2024). We return to the importance of family background later.

Our analysis also shows that, among tertiary-graduated individuals, there is little evidence of sorting into the first employer on the basis of EA-PGI. This is consistent with labor market entrants having limited information about firm and match quality, and with firms similarly knowing little about the productivity of new hires (Farber and Gibbons, 1996). Despite the absence of initial sorting, EA-PGI subsequently matter for transitions to higher-paying employers, suggesting that employer learning about workers' productivity and job mobility are important channels underlying the genetic gradient in income. Consistent with this, we observe steeper firm-quality trajectories among workers at the 90th EA-PGI percentile, while those at the 10th percentile exhibit flatter ones. While part of these results may reflect occupational sorting, the magnitude of firm-quality differences we observe suggests that employer characteristics, beyond occupation alone, play a crucial role, given the substantial variation in firm quality within occupations (e.g., Card, Heining, and Kline, 2013).

Because health is a well-documented determinant of educational and income disparities, we also examine whether it mediates the EA-PGI–income relationship. Although previous studies have shown that EA-PGI is associated with multiple diseases (Okbay et al., 2022), we find that the association between EA-PGI and cumulative disease burden is similar across education groups. This suggests that differences in health do not explain the stronger EA-PGI–income association among the tertiary-educated.

By Mendel's laws, parents of children with higher EA-PGI must have higher EA-PGI themselves, which results in higher parental education. This pattern is confirmed in our data: individuals in the 10th percentile of EA-PGI come from substantially lower socioeconomic backgrounds than those in the 90th percentile (Supplementary Table A.5). We therefore examine the role of parental EA-PGIs in accounting for the observed income gap across offspring's EA-PGI levels. Adjusting for parental EA-PGI reduced the offspring income gap by approximately 71 %, indicating that indirect effects play an important role in shaping income development.

Notably, most of this attenuation was driven by the father's EA-PGI rather than the mother's, a pattern not observed for educational attainment itself. This may reflect differential mechanisms: maternal resources, such as time and cognitive engagement, are particularly relevant for early-life development and educational choices, whereas paternal resources may become more influential during labor-market entry, through direct income support, occupational networks, or knowledge transmission about high-quality firms (Del Boca, Flinn, and Wiswall, 2014). These results are especially striking given Finland's relatively high level of gender equality in employment opportunities

(World Economic Forum, 2024).

This study advances the literature in several key ways. First, we analyse a substantially larger genotyped sample, including 12,871 parent–offspring trios, than previous studies, linked to decades of high-quality administrative data. This allows us to produce precise estimates of the genetic drivers of socioeconomic disparities. Second, we leverage continuous, rich information from tax records avoiding self-report bias, missing-not-at-random mechanisms, and measurement error that often affect survey-based studies. Third, by exploiting repeated measurements of workers and employers, and their links over time, we apply panel-data methods that account for worker and firm heterogeneity, which in turns allows us to provide novel evidence on the role of genetics in explaining the production of inequality in the labor markets. This allows us to decompose income variation into worker- and firm-level components, bridging genetics and modern labor economics (Card et al., 2018; Song et al., 2018; Kline, 2025).

Our study also has limitations. First, our sample consists solely of Finnish individuals who tend to be positively selected compared to the general population. This limits the generalizability to other ancestries or institutional settings. To address the fact that genotyped individuals are not fully representative of the Finnish population, we applied inverse-probability weighting to make the sample representative of all fresh graduates in the same calendar period (Davies et al., 2018). Weighted analyses produced results highly consistent with our main estimates (Supplementary Table A.11). Second, EA-PGI should not be interpreted as a pure "genetic endowment for educational attainment," as it may capture pathways correlated with educational attainment, such as lifestyle or social factors, that are modifiable and context-dependent. As such, changes in these underlying environments or social conditions could alter the observed relationship between EA-PGI and income. Lastly, as noted, the large attenuation we observe when controlling for parental EA-PGI indicates that indirect effects—particularly those operating through fathers—play an important role in shaping income differences. However, our setting does not allow us to disentangle the (parental) genetic and alternative environmental components underlying this result, which we leave to future work.

Taken together, our results indicate that EA-PGI is a robust predictor of income across the working life course, but only among tertiary-educated individuals. These effects operate primarily through individuals with higher EA-PGI experiencing faster and more frequent labor-market transitions to higher-quality employers, which in turn offer higher wages and thereby further amplify income inequality. Moreover, a substantial share of the EA-PGI–income association arises through indirect paternal genetic effects, suggesting that in the relatively recent generations studied (average graduation year 2000), fathers continued to play an important role in shaping offspring labor-market trajectories through untransmitted genetic factors.

Data and Methods

1.1 Data Sources

Genetic data. The genotyped sample was obtained from Finnish biobanks and consists of individuals who provided consent for research use of their blood samples. Participants in our dataset were drawn from several population-based epidemiological cohorts: 27,135 from the THL and 23,921 from the Blood Donor study.[1]

To quantify the genetic contribution to educational attainment, we constructed a polygenic index for years of education (EA-PGI) based on the largest genome-wide association study of educational attainment to date (Okbay et al., 2022), excluding Finnish cohorts to avoid overfitting. In our analysis sample, the PGI explained 7.1 %[2] of the variance in years of education (Supplementary Table A.12)—slightly lower than estimates from previous studies Lee, Wedow, Okbay, et al., 2018, possibly due to differences in cohort composition or educational classification.

Our sample tends to be positively selected (younger, better educated, with a higher share of women) compared to the population of fresh graduates. To analyze whether sample composition is a likely driver of our results, we apply an inverse probability weighting approach to make the sample representative of the graduates population (see e.g. Davies et al., 2018). Supplementary Table A.2 provides details on the re-weighting procedure and shows that it is effective in making the sample representative of the general population along the central dimensions that are initially unbalanced (including entry income, which we do not re-weight in our routine). Using these weights when estimating the income trajectories yields results that are qualitatively very similar to those in our main analysis, both when using the full genotyped sample and when using the family trios (Supplementary Table A.11).

Register data. We link the genotyped data to administrative registers from Statistics Finland (FOLK databases), covering the years 1987–2019. In addition, we utilize these register data independently, as they include the entire population of individuals permanently residing in Finland at the end of each year. The registers provide detailed information on employment histories, which allows us to identify the main employer at the end of each calendar year. They also contain income by source, from which we identify yearly labor income. We include people with zero income in the income analysis, thereby avoiding conditioning on employment. Given the long time series used in the analysis, all monetary values are deflated to 2010 EUR to account for inflation

[1] THL biobank website: `http://www.thl.fi/biobank`. Blood service biobank website: `https://www.veripalvelu.fi/en/biobank/`.

[2] Computed as incremental R^2, following Okbay et al. (2022), and controlling for gender fully interacted with year of birth indicators, biobank indicator and first 10 genetic principal components

and ensure comparability across time. The choice of the reference year is arbitrary; we choose 2010 because it lies midway between 2000 – the average graduation year in the sample – and the last follow-up year, 2019.

The registers further include demographic variables (gender, year of birth) and detailed educational information (highest degree, 4-digit field, school or institution ID, vocational vs. academic track), occupation, and industry codes.

Health registers. Health outcomes are obtained from two nationwide registers maintained by the Finnish Institute for Health and Welfare: the Care Register for Health Care (Hilmo) and the Register of Primary Health Care Visits (Avohilmo). In this study, Hilmo covers inpatient visits, operations, and specialized outpatient visits for the period to 1987–2024, when diagnoses follow ICD-9 and ICD-10 coding. Avohilmo, which uses ICD-10, covers primary care outpatient visits since 2011. For individuals absent from Hilmo, Avohilmo is used to complement the coverage. Both registers contain patient identifiers, care episode details, and one or more discharge diagnoses.

1.2 Methods

Polygenic Indices

We construct polygenic indices (PGIs) by aggregating single nucleotide polymorphisms (SNPs), common genetic variants identified in Genome-Wide Association Studies (GWAS) as predictive of years of education and health-related outcomes (see e.g. Biroli et al., 2025). SNPs are linearly combined using GWAS-derived effect sizes as weights, producing out-of-sample PGIs predictive of each trait of interest.

Our primary measure is the PGI for educational attainment (EA-PGI), standardized to mean zero and unit variance. Its distribution across education groups is shown in Supplementary Figure A.1. To control for ancestry and population stratification, we compute the first ten principal components of the genetic data and include them as covariates in all analyses.

Worker Ability and Firm Quality Measurement

To operationalize worker ability (or labor market productivity) θ_i, and firm quality ψ_J we estimate an Abowd, Kramarz, and Margolis (1999) (AKM) regression (Kline, 2024):

$$y_{it} = \mathbf{X}_{it}\beta + \psi_{J(i,t)} + \theta_i + \varepsilon_{it} \tag{1}$$

where y_{it} denotes log monthly labor income[3] for individual i in year t; $\mathbf{X}_{it}$ includes

[3] The outcome in (1) is monthly income, defined as annual earnings divided by number of months worked. Results are robust to using hourly wages derived from the Structure of Earnings Register (SES),

education fully interacted with calendar year and cubic age polynomial; $\psi_{J(i,t)}$ represents firm fixed effects; and θ_i are worker fixed effects. Estimated $\hat{\theta}_i$ provides a measure of worker productivity (unobserved heterogeneity) and $\hat{\psi}_{J(i,t)}$ offers firm-specific wage premia, interpreted as firm quality. Workers with higher fixed effects earn more across firms relative to a reference worker, holding observables constant. Similarly, firms with higher AKM fixed effects pay higher average wages, consistent with higher productivity serving as a proxy firm quality, as they capture persistent wage differences across firms after accounting for worker characteristics.

Except for the additive separability of firm and worker fixed effects, no functional form assumptions are made on either of the fixed effects, which are estimated nonparametrically. The estimation sample includes all full-time employees aged 20–60. The main employer is defined as the highest-paying ongoing job at year-end. To ensure labor market attachment, we restrict to workers earning at least 50% of the national median monthly income. the model is estimated by using the sample of firms linked by workers job-to-job transitions, also called connected set (Kline, 2025). Firms that do not experience hires or separations during their observation period are not part of the connected set and do not have an associated firm fixed effect.

For the estimation of the model, employment spells in very small firms (<5 employees) or shorter than four months are excluded. The resulting panel comprises 3.7 million workers, 31.6 million person-year observations, and 177.0 thousand firms. The AKM estimation is performed separately for two periods (1987–2003 and 2004–2019) due to computational constraints. Correlations of worker and firm fixed effects across the two periods are reasonably high (Supplementary Figure A.7).

We standardize worker and firm fixed effects to have zero mean and unit variance, and link them to the genotyped sample. Supplementary Tables A.3 and A.4 provide summary statistics for the sample used in estimating AKM, and standard statistics and income variance decomposition following AKM estimations.

Income and Health Trajectory Model

To study how genetic predispositions affect income over the life cycle, we estimate:

$$y_{icmt} = \alpha + \tau_c + \tau_m + \beta_t PGI_i + \gamma X_i + \varepsilon_{icmt} \qquad (2)$$

where y_{icmt} is labor income of individual i or the Charlson Comorbidity Index (see below), in birth cohort c, calendar year m, and number of years since graduation t. PGI_i is standardised EA-PGI; τ_c and τ_m are cohort and year effects; and X_i includes ten genetic principal components, gender, and biobank indicator (THL or Blood Donors).

which covers the whole public sector and a sample of about half of the private sector. To maximize sample size and coverage, we use monthly earnings as our baseline income measure.

The coefficients of interest are β_t, which capture the income differential for EA-PGI over time since graduation. Since PGIs are randomly assigned at conception, β_t has a causal interpretation, conditional on adequate control for population stratification via X_i. Any residual stratification would lead the coefficients to capture a compound effect of own genetics and other environmental factors. While assigning sign to the bias is not immediate, we believe that it is reasonable to consider our estimates as conservative lower bounds of true causal effects of own EA-PGI.[4]

Our model income trajectories specifications include continue EA-PGI evaluated at 10th and 90th percentiles of EA-PGI distribution. The results are similar when specifying EA-PGI via deciles (Supplementary Figure A.2). Finally, when computing cumulated lifetime income we discount income by a 3% interest rate, computing its present value upon graduation. In the computation, we sum over all discounted income rows (including zeros) between graduation year and up to 25 years later.

Income decomposition by EA-PGI group and over time since graduation

We implement the approach by Hahn, Hyatt, and Janicki (2021) to decompose the log-income growth separately by EA-PGI group (1st and 10th deciles).

Each year since graduation t, workers are partitioned into one of four groups: *stayers* (workers who stay with the same employer); *employer-to-employer transitions* (workers who change firm); *entrants from non-employment* (hires from nonemployment); *exiters to non-employment* (incumbent workers separating to nonemployment). The average income growth between $t-1$ and t is decomposed into four weighted contributions based on the four worker types (weighted by the share of workers in each worker type). Since entrants and exiters move between employment and non-employment, their contribution is obtained by comparing their average income to that of the workers who are continuously employed in the time period.

In line with Hahn, Hyatt, and Janicki (2021), and confirmed by our analysis, job-to-job movers' transitions are associated with large earnings gains for individuals. Moreover, the entrants from nonemployment earn substantially less than the continuously-employed workers to which their salary is compared to. Hence, their entry into employment lessens (subtracts from) the average earnings and their contribution to the average earnings growth is negative. The opposite occurs for exiters to nonemployment: they also tend to earn less than the continuously-employed workers, but because these low-paying jobs dissolve, this contributes positively to the earnings growth.

We present results by cumulating the income growth components over $t = 1, ..., 25$.

[4]The PGI captures only the contribution of common genetic variants identified in external GWAS, not the full genetic architecture of education.

Charlson Comorbidity Index

The Charlson Comorbidity Index (CCI) (Charlson et al., 1987; Deyo, Cherkin, and Ciol, 1992) assigns fixed weights to comorbid conditions associated with higher mortality risk. The index is the weighted sum of an individual's comorbidities, with weights derived from Cox regression models.

We compute CCI scores using the ICCI R package (Detrois, 2024), which implements the ICD-9 and ICD-10 coding (Quan et al., 2005) via the comorbidity package by Gasparini (2018).[5] For each individual, we compute cumulative CCI scores by age, recalculating the index at successive cutoffs (0–19, 0–20, . . . , up to 0–50 years).

Definitions and Sample Restrictions. For the trajectory-based analyses, we construct a panel of all genotyped adults with either secondary or tertiary qualification, followed from year of graduation onward between 1987 and 2019. To ensure that individuals in our sample have completed their education phase, we remove those with secondary degree that have not been observed past age 30; people that obtain tertiary degree before age 30 are retained in the sample. The final sample includes 51,056 individuals and 963,715 person-year observations (see Supplementary Table A.1).

Acknowledgments

We would like to thank Adrian Adermon, Mikko Arvas, Pietro Biroli, Rosa Cheesman, Lena Hensvik, Paul Hufe, Andrea Ichino, Francis Kramarz, Martin Nybom, Gerard van den Berg, and participants to the European Social Science Genetics Network Conference 2023 (Bologna), Nordic Summer Institute 2025 (Uppsala), Workshop on Inequality and education 2025 (Modena), Workshop in Wages, Employment, and Inequality 2025 (Helsinki), for useful comments and discussions.

We thank all study participants for their generous participation in biobank research, and we are grateful to Blood Service Biobank and to THL Biobank for issuing the data permits to utilize the genetic data (BSB permit no. 005-2020 and THLBB2023_47). The project fall under IRB ethical approval HUS/1311/2020, Statistics Finland data permit TK/125/07.03.00/2020, and Findata permit THL/4014/14.02.00/2022. Lombardi gratefully acknowledges funding from the Research Council of Finland (decision no. 350399) and Yrjö Jahnsson Foundation (decision no. 20247803).

[5]The package accommodates multiple ICD versions. Source code: ICCI GitHub repository.

References

Abdellaoui, Abdel, Loic Yengo, Karin JH Verweij, and Peter M Visscher (2023). "15 years of GWAS discovery: realizing the promise". *The American Journal of Human Genetics* 110.2, 179–194.

Abowd, John M., Francis Kramarz, and David N. Margolis (1999). "High wage workers and high wage firms". *Econometrica* 67.2, 251–333.

Agardh, Emilie, Peter Allebeck, Johan Hallqvist, Tahereh Moradi, and Anna Sidorchuk (2011). "Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis". *International Journal of Epidemiology* 40.3, 804–818.

Akimova, Evelina T, Tobias Wolfram, Xuejie Ding, Felix C Tropf, and Melinda C Mills (2025). "Polygenic prediction of occupational status GWAS elucidates genetic and environmental interplay in intergenerational transmission, careers and health in UK Biobank". *Nature Human Behaviour* 9.2, 391–405.

Barth, Daniel, Nicholas W Papageorge, and Kevin Thom (2020). "Genetic Endowments and Wealth Inequality.pdf". *Journal of Political Economy* 128.4, 1474–1522.

Biroli, Pietro, Titus Galama, Stephanie von Hinke, Hans Van Kippersluis, Cornelius A Rietveld, and Kevin Thom (2025). "The Economics and econometrics of gene–environment interplay". *Review of Economic Studies*, rdaf034.

Black, Sandra E, Paul J Devereux, Fanny Landaud, and Kjell G Salvanes (2025). "The (un) importance of inheritance". *Journal of the European Economic Association* 23.3, 1060–1094.

Bundy, Joshua D, Katherine T Mills, Hua He, Thomas A LaVeist, Keith C Ferdinand, Jing Chen, and Jiang He (2023). "Social determinants of health and premature death among adults in the USA from 1999 to 2018: a national cohort study". *The Lancet Public Health* 8.6, e422–e431.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018). "Firms and Labor Market Inequality: Evidence and Some Theory". *Journal of Labor Economics* 36 (S1), S13–S70.

Card, David, Jörg Heining, and Patrick Kline (2013). "Workplace heterogeneity and the rise of West German wage inequality". *The Quarterly journal of economics* 128.3, 967–1015.

Carvalho, Leandro S (2025). "Genetics and Socioeconomic Status: Some Preliminary Evidence on Mechanisms". *Journal of political economy microeconomics* 3.3, 429–476.

Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie (1987). "A new method of classifying prognostic comorbidity in longitudinal studies: development and validation". *Journal of Chronic Diseases* 40.5, 373–383.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler (2016). "The association between income and life expectancy in the United States, 2001-2014". *Jama* 315.16, 1750–1766.

Cunha, Flavio and James Heckman (2007). "The Technology of Skill Formation". *American Economic Review* 97.2, 31–47.

Davies, Neil M., Matt Dickson, George Davey Smith, Gerard J. Van Den Berg, and Frank Windmeijer (2018). "The causal effects of education on health outcomes in the UK Biobank". *Nature Human Behaviour* 2.2, 117–125.

Del Boca, Daniela, Christopher Flinn, and Matthew Wiswall (2014). "Household choices and child development". *Review of Economic Studies* 81.1, 137–185.

Detrois, Kira E. (2024). *ICCI: Calculates the Charlson Comorbidity Index*. R package version 2.3.1.

Deyo, R. A., D. C. Cherkin, and M. A. Ciol (1992). “Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases”. *Journal of Clinical Epidemiology* 45.6, 613–619.

Dwyer-Lindgren, Laura, Parkes Kendrick, Yekaterina O Kelly, Dillon O Sylte, Chris Schmidt, Brigette F Blacker, Farah Daoud, Amal A Abdi, Mathew Baumann, Farah Mouhanna, et al. (2022). “Life expectancy by county, race, and ethnicity in the USA, 2000–19: a systematic analysis of health disparities”. *The Lancet* 400.10345, 25–38.

Farber, Henry S and Robert Gibbons (1996). “Learning and wage dynamics”. *The Quarterly Journal of Economics* 111.4, 1007–1047.

Gasparini, Alessandro (2018). “comorbidity: An R package for computing comorbidity scores”. *Journal of Open Source Software* 3.23, 648.

Hahn, Joyce K., Henry R. Hyatt, and Hubert P. Janicki (2021). “Job ladders and growth in earnings, hours, and wages”. *European Economic Review* 133, 103654.

Harden, K. Paige and Philipp D. Koellinger (2020). “Using genetics for social science”. *Nature Human Behaviour* 4.6, 567–576.

Ichino, Andrea, Aldo Rustichini, and Giulio Zanella (2024). “College, cognitive ability, and socioeconomic disadvantage: policy lessons from the UK in 1960-2004”. *Review of Economic Studies (forthcoming)*.

Kline, Patrick (2024). “Firm wage effects”. *Handbook of Labor Economics* 5, 115–181.

Kline, Patrick M (2025). “Firm wage effects”. NBER Working Paper 33084.

Kong, Augustine, Gudmar Thorleifsson, Michael L Frigge, Bjarni J Vilhjalmsson, Alexander I Young, Thorgeir E Thorgeirsson, Stefania Benonisdottir, Asmundur Oddsson, Bjarni V Halldorsson, Gisli Masson, et al. (2018). “The nature of nurture: Effects of parental genotypes”. *Science* 359.6374, 424–428.

Lee, James J., Robbee Wedow, Aysu Okbay, et al. (2018). “Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals”. *Nature Genetics* 50.8. Published online 23 Jul 2018, 1112–1121.

Lynch, John W., George Davey Smith, George A. Kaplan, and James S. House (2000). “Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions”. *BMJ* 320.7243, 1200–1204.

Martschenko, Daphne, Sam Trejo, and Benjamin W Domingue (2019). “Genetics and education: Recent developments in the context of an ugly history and an uncertain future”. *AERA Open* 5.1, 2332858418810516.

Mills, Melinda C (2022). “Sociogenomics: theoretical and empirical challenges of integrating molecular genetics into sociological thinking”. *Handbook of sociological science*. Edward Elgar Publishing, 250–270.

Newman, David B., Amie M. Gordon, and Wendy Berry Mendes (2025). “Income and education show distinct links to health and happiness in daily life”. *Nature Human Behaviour*. Published online 8 Aug 2025.

Okbay, Aysu et al. (2022). “Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals”. *Nature Genetics* 54.4, 437–449.

Pallesen, Anna Vera Jørring, Jochen O. Mierau, Frederik Kølby Christensen, and Laust Hvas Mortensen (2024). "Educational and income inequalities across diseases in Denmark: a register-based cohort study". *The Lancet Public Health* 9.11, e916–e924.

Pickett, Kate E. and Richard G. Wilkinson (2015). "Income inequality and health: A causal review". *Social Science & Medicine* 128, 316–326.

Quan, Hude, Victoria Sundararajan, Paul Halfon, Andrew Fong, Bernard Burnand, Jean-Christophe Luthi, Lawrence D. Saunders, Cynthia A. Beck, Lorraine Feasby, and William A. Ghali (2005). "Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data". *Medical Care* 43.11, 1130–1139.

Roemer, John E. and Alain Trannoy (2016). "Equality of Opportunity: Theory and Measurement". *Journal of Economic Literature* 54.4, 1288–1332.

Rustichini, Aldo, William G Iacono, James J Lee, and Matt McGue (2023). "Educational attainment and intergenerational mobility: A polygenic score analysis". *Journal of Political Economy* 131.10, 2724–2779.

Schwandt, Hannes, Janet Currie, Marlies Bär, James Banks, Paola Bertoli, Aline Bütikofer, Sarah Cattan, Beatrice Zong Ying Chao, Claudia Costa, Libertad González, Veronica Grembi, Kristiina Huttunen, René Karadakic, Lucy Kraftman, Sonya Krutikova, Stefano Lombardi, Peter Redler, Carlos Riumallo-Herl, Ana Rodríguez-González, Kjell G. Salvanes, Paula Santana, Josselin Thuilliez, Eddy van Doorslaer, Tom van Ourti, Joachim K. Winter, Bram Wouterse, and Amelie Wuppermann (2021). "Inequality in mortality between Black and White Americans by age, place, and cause, and in comparison to Europe, 1990–2018". *Proceedings of the National Academy of Sciences of the United States of America* 118.40. Published online 28 Sept 2021, e2104684118.

Schwandt, Hannes and Till Von Wachter (2020). "Socioeconomic decline and death: Midlife impacts of graduating in a recession".

Sjaarda, Jennifer and Zoltán Kutalik (2023). "Partner choice, confounding and trait convergence all contribute to phenotypic partner similarity". *Nature human behaviour* 7.5, 776–789.

Sohail, Mashaal, Robert M Maier, Andrea Ganna, Alex Bloemendal, Alicia R Martin, Michael C Turchin, Charleston WK Chiang, Joel Hirschhorn, Mark J Daly, Nick Patterson, et al. (2019). "Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies". *Elife* 8, e39702.

Song, Jae, David J Price, Fatih Guvenen, and Nicholas Bloom (2018). "Firming up inequality". *The Quarterly Journal of Economics* 134.1, 1–50.

Tan, Tammy, Hariharan Jayashankar, Junming Guan, Seyed Moeen Nehzati, Mahdi Mir, Michael Bennett, Esben Agerbo, Rafael Ahlskog, Ville Pinto de Andrade Anapaz, Bjørn Olav Åsvold, Stefania Benonisdottir, Laxmi Bhatta, Dorret I. Boomsma, Ben Brumpton, Archie Campbell, Christopher F. Chabris, Rosa Cheesman, Zhengming Chen, Eco de Geus, Erik A. Ehli, Abdelrahman G. Elnahas, Estonian Biobank Research Team, Finngen Authors, Andrea Ganna, Alexandros Giannelis, Liisa Hakaste, Ailin Falkmo Hansen, Alexandra Havdahl, Caroline Hayward, Jouke-Jan Hottenga, Mikkel Aagaard Houmark, Kristian Hveem, Jaakko Kaprio, Arnulf Langhammer, Antti Latvala, James J. Lee, Mikko Lehtovirta, Liming Li, Kuang Lin, Richard Karlsson Linner, Stefano Lombardi, Nicholas G. Martin, Matt McGue, Sarah E. Medland, Andres Metspalu, Brittany L. Mitchell, Guiyan Ni, Ilja M. Nolte, Matthew T. Oetjens, Sven Oskarsson, Teemu Palviainen, Rashmi B. Prasad, Anu Reigo, Kadri Reis, Julia Sidorenko, Karri Silventoinen, Harold Snieder, Tiinamaija Tuomi, Bjarni J. Vilhjalmsson, Robin G. Walters, Emily A. Willoughby, Bendik S. Winsvold, Eivind Ystrom, Jonathan Flint, Loic Yengo, Peter M. Visscher, Augustine Kong, Elliot M. Tucker-Drob, Richard Border, David Cesarini, Patrick Turley, Aysu Okbay, Daniel J.

Benjamin, and Alexander Strudwick Young (2024). "Family-GWAS reveals effects of environment and mating on genetic associations". *medRxiv*. preprint.

Tillmann, T., J. Vaucher, A. Okbay, H. Pikhart, A. Peasey, R. Kubinova, A. Pajak, A. Tamosiunas, S. Malyutina, F. P. Hartwig, K. Fischer, G. Veronesi, T. Palmer, J. Bowden, G. Davey Smith, M. Bobak, and M. V. Holmes (2017). "Education and coronary heart disease: mendelian randomisation study". *BMJ (Clinical research ed.)* 358, j3542.

Vaccarella, Salvatore, Damien Georges, Freddie Bray, Ophira Ginsburg, Hadrien Charvat, Pekka Martikainen, Henrik Brønnum-Hansen, Patrick Deboosere, Matthias Bopp, Mall Leinsalu, Barbara Artnik, Valentina Lorenzoni, Esther de Vries, Michael Marmot, Paolo Vineis, Johan Mackenbach, and Wilma Nusselder (2023). "Socioeconomic inequalities in cancer mortality between and within countries in Europe: a population-based study". *The Lancet Regional Health - Europe* 25, 100551.

Wang, Biyao, Jessie R Baldwin, Tabea Schoeler, Rosa Cheesman, Wikus Barkhuizen, Frank Dudbridge, David Bann, Tim T Morris, and Jean-Baptiste Pingault (2021). "Robust genetic nurture effects on education: A systematic review and meta-analysis based on 38,654 families across 8 cohorts". *The American Journal of Human Genetics* 108.9, 1780–1791.

World Economic Forum, WEF (2024). "Global Gender Gap Report 2024". Tech. rep. Accessed: 2025-11-12. World Economic Forum.

Young, Alexander I., S. M. Nehzati, Snaedis Benonisdottir, et al. (2022). "Mendelian imputation of parental genotypes improves estimates of direct genetic effects". *Nature Genetics* 54.6, 897–905.

A Supplementary Figures and Tables

A.1 Supplementary Figures

Figure A.1: Density of EA-PGI by highest education level and track
The density plot uses full analysis sample with non-missing education track information ($N = 50,940$). The blue lines correspond to secondary, and the red - to tertiary degree level. Solid lines correspond to academic, and the dashed lines - to vocational education tracks.

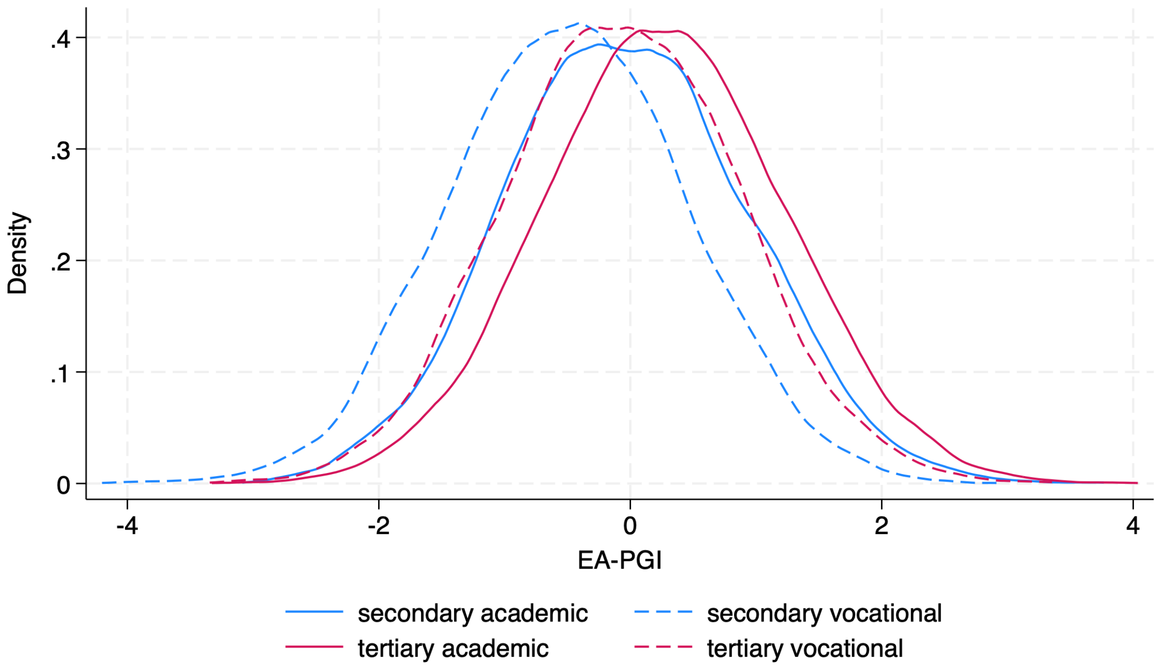


Figure A.2: Average annual income of tertiary-educated individuals by EA-PGI level, over time

The figure uses a subset of tertiary-educated individuals ($N = 32{,}364$). The lines correspond to ten deciles of the EA-PGI distribution. Average income estimated from a regression of annual income on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95 % CIs.

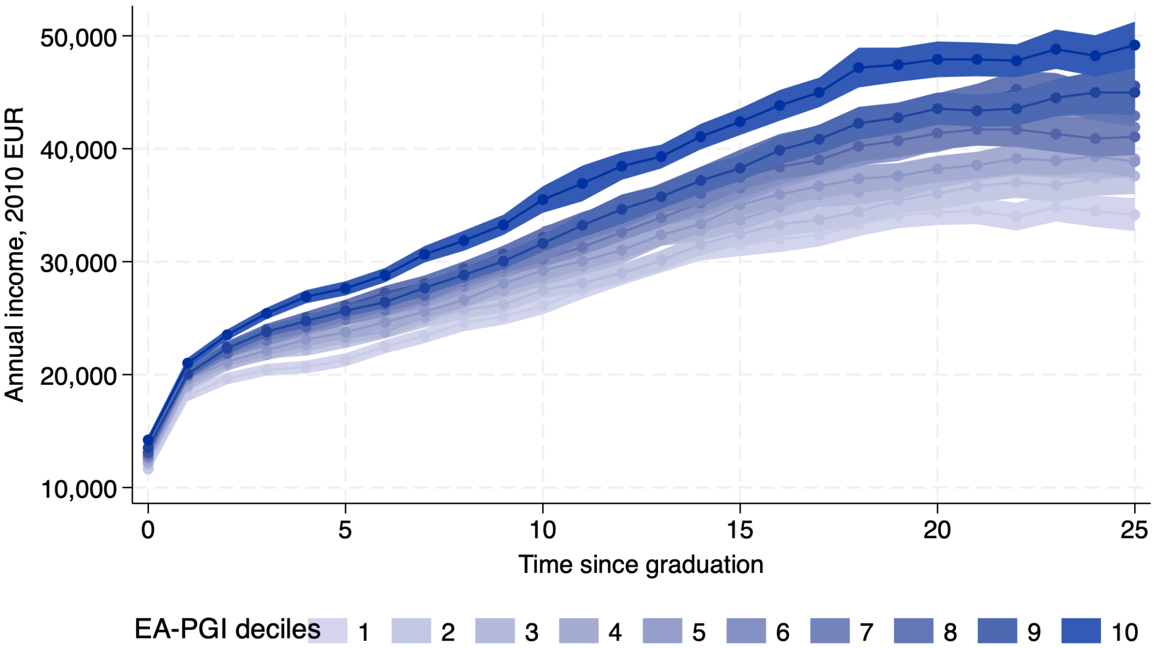
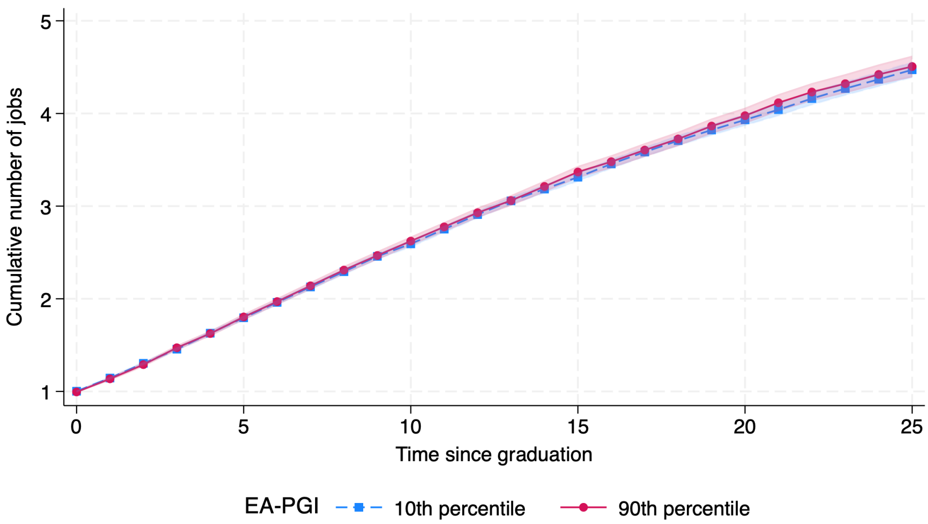


Figure A.3: Employer mobility and quality of secondary-educated individuals by EA-PGI level, over time

Panel (a) plots the average number of employment spells, and Panel (b) - the average firm quality index over time since graduation. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Both panels are estimated from a regression of respective outcome on EA-PGI fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The estimation sample is restricted to secondary-educated workers with non-missing firm identifiers ($N = 17{,}432$) in panel (a) and non-missing firm quality index ($N = 15{,}579$) in panel (b). The shaded areas correspond to 95 % CIs.

(a) Number of jobs since labor market entry



(b) Firm quality since labor market entry

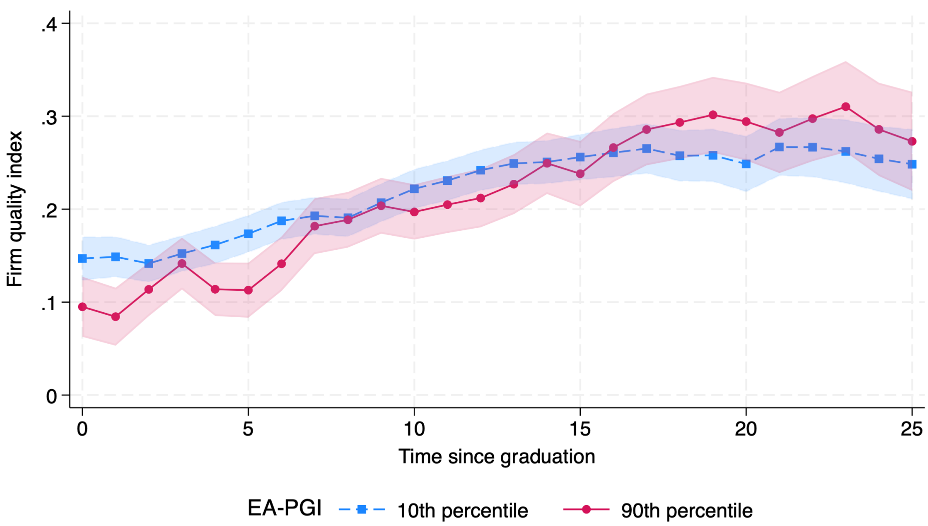
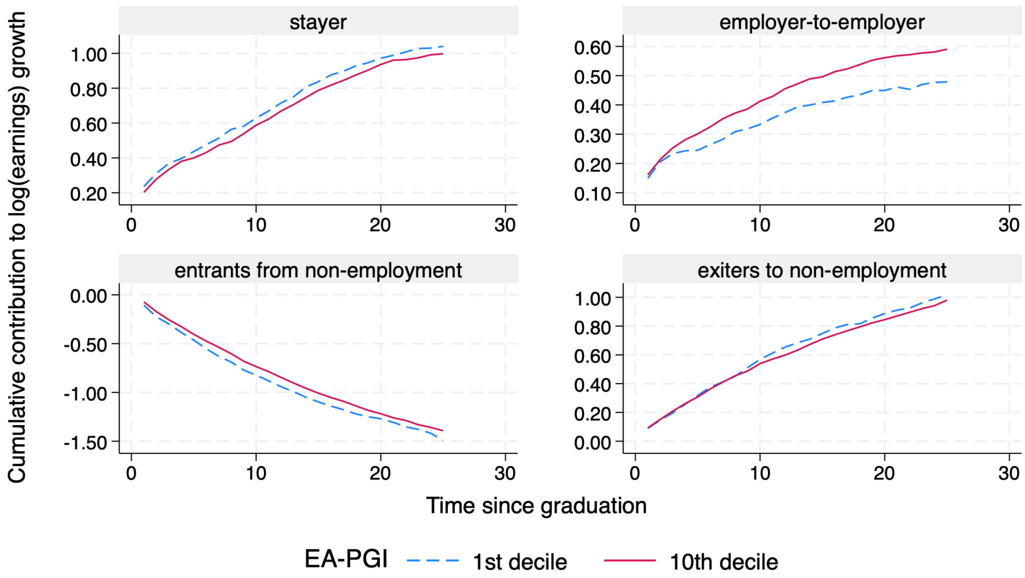


Figure A.4: Decomposition of cumulative growth in log annual income of tertiary-educated individuals by EA-PGI level, over time, including mobility into and out of non-employment

Panel (a) reports cumulative contributions of four mobility types to overall cumulative log earnings growth over time since graduation. Panel (b) reports average employment shares of four mobility types over time since graduation. The blue line corresponds to 1st and the red - to 10th decile of EA-PGI distribution. The decomposition follows the algorithm in Hahn, Hyatt, and Janicki, 2021. The sample for the decomposition is restricted to tertiary-educated workers ($N = 32{,}364$).

(a) Cumulative contribution to log(earnings) growth



(b) Employment share

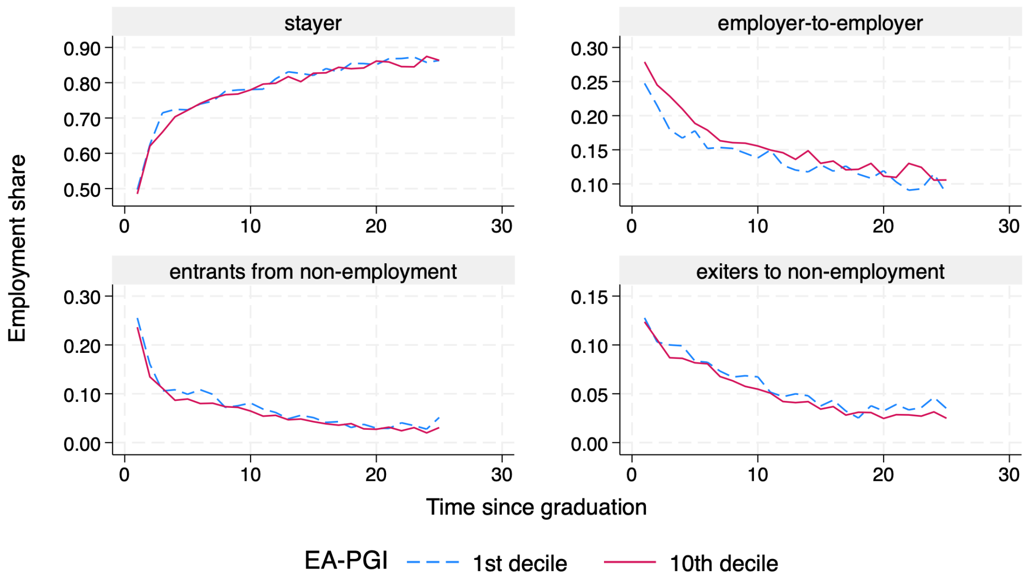


Figure A.5: Average health index by EA-PGI level, over time and by education, conditional on parental EA-PGI among parent-offspring trios

Panel **(a)** uses full sample of parent-offspring sample with non-missing Charlson Comorbidity Index ($N = 50{,}963$), while Panels **(b)** and **(c)** use subset of these workers based on their highest qualification being either secondary ($N = 18{,}669$) or tertiary degree ($N = 32{,}294$), respectively. The blue line corresponds to 10th and the red - to 90th percentile of EA-PGI distribution. Average health index estimated from a regression of Charlson Comorbidity Index on the set of own, maternal and paternal EA-PGI, all fully interacted with indicators measuring years since graduation, and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95 % CIs.

(a) Pooled **(b) Secondary** **(c) Tertiary**

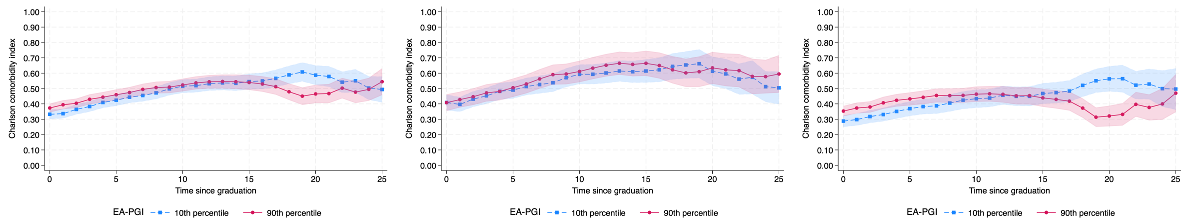


Figure A.6: Average annual income of tertiary-educated individuals by own or parental EA-PGI levels, over time

The figure uses a subset of parent-offspring trios with highest offspring qualification being tertiary degree ($N = 7{,}808$). The lines in Panel **(a)** correspond to 10th and 90th percentiles of offspring EA-PGI; in Panel **(b)** - 10th and 90th percentiles of paternal EA-PGI; and in Panel **(c)** - 10th and 90th percentiles of maternal EA-PGI. Average income estimated from a regression of annual income on the set of own, maternal and paternal EA-PGI, all fully interacted with indicators measuring years since graduation, and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95 % CIs.

(a) Own EA-PGI **(b) Father EA-PGI** **(c) Mother EA-PGI**

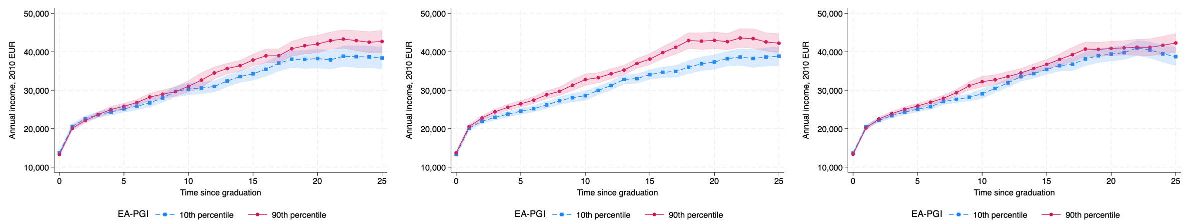
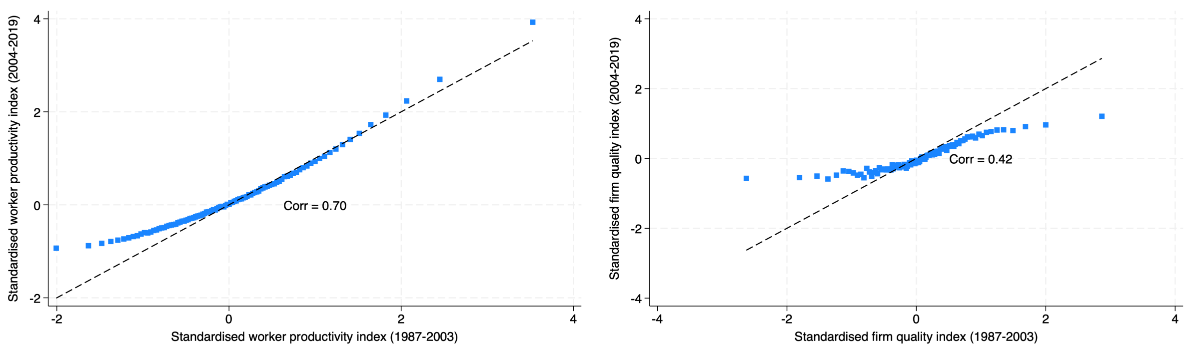


Figure A.7: AKM worker productivity and firm quality indices between two AKM estimation periods (1987-2003 and 2004-2019)

Panel **(a)** is a binscatter plot of worker productivity indices estimated using matched employer-employee data between 2004-2019 (on the y axis) and 1987-2003 (on the x axis). Panel **(b)** is a binscatter plot of firm quality indices estimated using matched employer-employee data between 2004-2019 (on the y axis) and 1987-2003 (on the x axis). The dashed black lines correspond to 45° line. The sample in Panel **(a)** are individuals with both worker productivity indices non-missing ($N = 13,715$); in Panel **(b)** - individuals with both firm quality indices non-missing ($N = 32,494$).

(a) Worker productivity indices **(b) Firm quality indices**



A.2 Supplementary Tables

Table A.1: Working sample in trajectory analysis

The table reports sample counts by working sample restrictions. Columns (1)-(3) report person-year observation counts in full sample and by biobanks, respectively. Columns (4)-(6) report unique individual counts in full sample and by biobanks.

	Person-year observations			Unique individuals		
	All (1)	THL (2)	BDB (3)	All (4)	THL (5)	BDB (6)
Start	5,374,521	3,963,254	1,411,267	176,523	132,171	44,352
Keep individuals with non-missing genetic PCs	5,335,564	3,963,254	1,372,310	175,050	132,171	42,879
Keep graduates only	3,215,453	1,957,308	1,258,145	98,810	59,416	39,394
Keep graduates with non-missing graduation year	3,215,453	1,957,308	1,258,145	98,810	59,416	39,394
Graduated between 1970 and 2020	3,139,889	1,882,118	1,257,771	96,186	56,803	39,383
Observed between 0 and 25 years since graduation	1,692,473	1,065,113	627,360	96,166	56,786	39,380
Followed from 0 years since graduation	1,000,872	576,894	423,978	57,956	29,205	28,751
Followed at least up to age 30 (if secondary)	963,715	567,236	396,479	51,056	27,135	23,921

Table A.2: Average population and sample characteristics of graduates
Panel **(a)** compares full analysis sample ($N = 51{,}056$) and Panel **(b)** compares subsample of parent-offspring trios ($N = 12{,}871$) to the population of fresh graduates. The population of fresh graduates is selected from the full Finnish population with same criteria described in Table A.1, except for conditioning on genotypes and genetic principal components. Column (1) reports average characteristics in the population of fresh graduates. Columns (2) and (3) report average characteristics in the corresponding sample, before and after applying inverse probability weights, respectively. Columns (4) and (5) report p-values for the equality of means between population and working sample, before and after applying inverse probability weights, respectively. All p-values are adjusted for multiple hypotheses testing using Holm correction. Columns (6) and (7) report population and sample counts, respectively. Inverse probability weights are estimated to balance the sample according to year of birth and graduation year (fully interacted with highest education level and gender), and rural area indicator fully interacted with gender.

(a) Genotyped sample

	Population (1)	Genotyped sample (2)	Reweighted sample (3)	$\Delta^{(1)}_{(2)}$ (4)	$\Delta^{(1)}_{(3)}$ (5)	$N^{(1)}$ (6)	$N^{(2)}$ (7)
Cohort: 1950-59	0.01	0.01	0.01	0.000	1.000	1,599,332	51,056
Cohort: 1960-69	0.17	0.22	0.16	0.000	1.000	1,599,332	51,056
Cohort: 1970-79	0.34	0.36	0.34	0.000	1.000	1,599,332	51,056
Cohort: 1980-89	0.36	0.29	0.37	0.000	1.000	1,599,332	51,056
Cohort: 1990-99	0.13	0.11	0.13	0.000	1.000	1,599,332	51,056
Graduation age: 16-20	0.36	0.31	0.36	0.000	1.000	1,599,332	51,056
Graduation age: 21-25	0.39	0.43	0.39	0.000	1.000	1,599,332	51,056
Graduation age: 26-30	0.25	0.25	0.24	0.001	1.000	1,599,332	51,056
Education: secondary	0.44	0.37	0.44	0.000	1.000	1,599,332	51,056
Education: tertiary	0.56	0.63	0.56	0.000	1.000	1,599,332	51,056
Male	0.48	0.39	0.48	0.000	1.000	1,599,332	51,056
Married	0.10	0.13	0.11	0.000	0.000	1,599,332	51,056
Rural	0.24	0.25	0.25	0.712	1.000	1,599,332	51,056
Cumulated income	279,534	302,077	277,373	0.000	0.692	1,599,332	51,056
Income at t=0	9,301	9,527	9,328	0.000	1.000	1,599,332	51,056

(b) Family trio sample

	Population (1)	Genotyped family trios (2)	Reweighted family trios (3)	$\Delta^{(1)}_{(2)}$ (4)	$\Delta^{(1)}_{(3)}$ (5)	$N^{(1)}$ (6)	$N^{(2)}$ (7)
Cohort: 1950-59	0.01	0.01	0.01	0.446	1.000	1,599,332	12,871
Cohort: 1960-69	0.17	0.15	0.17	0.001	1.000	1,599,332	12,871
Cohort: 1970-79	0.34	0.35	0.34	0.017	1.000	1,599,332	12,871
Cohort: 1980-89	0.36	0.38	0.36	0.000	1.000	1,599,332	12,871
Cohort: 1990-99	0.13	0.11	0.12	0.000	1.000	1,599,332	12,871
Graduation age: 16-20	0.36	0.34	0.36	0.000	1.000	1,599,332	12,871
Graduation age: 21-25	0.39	0.42	0.40	0.000	1.000	1,599,332	12,871
Graduation age: 26-30	0.25	0.24	0.24	0.086	1.000	1,599,332	12,871
Education: secondary	0.44	0.39	0.44	0.000	1.000	1,599,332	12,871
Education: tertiary	0.56	0.61	0.56	0.000	1.000	1,599,332	12,871
Male	0.48	0.40	0.48	0.000	1.000	1,599,332	12,871
Married	0.10	0.12	0.12	0.000	1.000	1,599,332	12,871
Rural	0.24	0.27	0.24	0.000	1.000	1,599,332	12,871
Cumulated income	279,534	282,107	283,464	0.446	1.000	1,599,332	12,871
Income at t=0	9,301	9,782	9,557	0.000	1.000	1,599,332	12,871

Table A.3: Descriptive statistics in AKM sample
The table reports descriptive statistics in full matched employee-employer panel used in AKM estimation. Column (1) reports sample mean, column (2) - standard deviation, and column (3) - person-year observation counts for the variables in rows.

	Mean (1)	SD (2)	N (3)
Cohort: 1946-1955	0.231	0.421	7,225,843
Cohort: 1956-1965	0.299	0.458	9,379,633
Cohort: 1966-1975	0.240	0.427	7,504,156
Cohort: 1976-1985	0.162	0.369	5,086,257
Cohort: 1986-1995	0.064	0.245	2,010,473
Cohort: 1996-2005	0.004	0.062	121,624
Age group: 20-29	0.200	0.400	6,875,154
Age group: 30-39	0.288	0.453	9,929,190
Age group: 40-49	0.288	0.453	9,926,283
Age group: 50-59	0.212	0.409	7,306,357
Age group: 60-69	0.011	0.106	391,987
Male	0.604	0.489	34,428,971
Education level: Compulsory	0.212	0.409	7,314,987
Education level: Secondary	0.447	0.497	15,406,896
Education level: Tertiary	0.340	0.474	11,707,088
Firm size	1,690	4,283	34,428,971
Annual total earning, 2015 EUR	27,319	30,515	34,428,936
Months worked in a year	11.680	1.241	34,428,971
Monthly total earning, 2015 EUR	2,330	2,585	34,428,936
Firm quality index	0.365	0.771	15,699,075
Worker productivity index	0.122	0.990	16,162,338

Table A.4: AKM summary statistics and variance decomposition

The table reports summary statistics and variance decomposition following AKM estimations. The dependent variable in AKM estimation is log monthly earnings calculated as the ratio of total annual earnings by number of months worked. The sample includes employees aged between 20 and 60, with monthly earnings above 50% of national median monthly income, working in firms with at least 5 employees and for at least four months in a calendar year. The estimations control for calendar year indicators, education level, cubic polynomial in age, as well as interactions of calendar year and age polynomial with education level. For further details, see Section 1.2.

	Dependent variable: log monthly earnings	
	1987-2003	2004-2019
Standard deviation of outcome	0.5003	0.4614
N largest connected set	16 862 428	15 435 023
N singletons	275 680	374 028
N estimation sample	16 586 748	15 060 995
Panel A: Summary of parameter estimates		
N worker FE	1 881 715	1 842 564
N firm FE	126 605	50 430
Std. dev. of worker FE	0.2969	0.3208
Std. dev. of firm FE	0.1067	0.1027
Std. dev. of Xb	0.3416	0.2437
Std. dev. of residual	0.1587	0.1561
Corr(worker FE, firm FE)	0.1054	0.2496
RMSE	0.1693	0.1669
Adjusted R2	0.8846	0.8681
Panel B: Share of outcome variance attributed to		
Worker FE	0.3547	0.4868
Firm FE	0.0458	0.0499
Cov(worker FE, firm FE)	0.0269	0.0778
Xb and associated covariances	0.4712	0.2703
Residual	0.1014	0.1153

Table A.5: Descriptive statistics by EA-PGI level

The table reports descriptive statistics among tertiary-educated individuals ($N = 32{,}364$) by deciles of EA-PGI. Columns (1)-(3) report sample means in 1st, 2nd-9th and 10th decile of EA-PGI, respectively. Column (4) reports difference in sample means and standard error of the difference in parentheses for the 2nd-9th decile relative to 1st decile, respectively. Column (5) reports difference in sample means and standard error of the difference in parentheses for the 10th decile relative to 1st decile, respectively. Column (6) reports total sample count for each variable considered. All estimates control for first ten genetic principal components.

	Sample means			Diff.				
	1st decile	2nd-9th deciles	10th decile	2nd-9th deciles		10th decile		
	(1)	(2)	(3)	(4)		(5)		(6)
Male	0.267	0.341	0.402	0.074***	(0.011)	0.135***	(0.013)	32,364
Birth year	1977.2	1977.3	1977.6	0.170	(0.231)	0.455	(0.272)	32,364
Mother edu: compulsory	0.332	0.250	0.175	-0.082***	(0.010)	-0.157***	(0.012)	32,364
Mother edu: secondary	0.413	0.355	0.258	-0.059***	(0.011)	-0.155***	(0.013)	32,364
Mother edu: tertiary	0.247	0.386	0.554	0.139***	(0.011)	0.307***	(0.013)	32,364
Mother edu: missing	0.008	0.009	0.012	0.001	(0.002)	0.005	(0.003)	32,364
Father edu: compulsory	0.351	0.278	0.196	-0.073***	(0.010)	-0.155***	(0.012)	32,364
Father edu: secondary	0.388	0.314	0.218	-0.074***	(0.011)	-0.170***	(0.012)	32,364
Father edu: tertiary	0.220	0.366	0.551	0.147***	(0.011)	0.332***	(0.013)	32,364
Father edu: missing	0.042	0.042	0.035	0.001	(0.005)	-0.007	(0.005)	32,364
Age at graduation	24.506	24.854	25.245	0.349***	(0.054)	0.740***	(0.063)	32,364
Graduation: years since predicted graduation	2.256	2.211	2.126	-0.045	(0.049)	-0.131*	(0.058)	32,364
Age at first job	26.332	26.448	26.739	0.116	(0.072)	0.407***	(0.085)	31,015
First job: years since predicted graduation	3.588	3.302	3.092	-0.286***	(0.071)	-0.496***	(0.084)	31,015
Annual average income at first job	11,729	12,811	14,208	1,082***	(242)	2,478***	(285)	32,364
AKM firm FE of first job	0.217	0.236	0.268	0.019	(0.022)	0.052*	(0.026)	16,860
AKM firm FE at t=15	0.407	0.462	0.527	0.055	(0.035)	0.120**	(0.043)	8,004

Table A.6: Worker productivity and EA-PGI
The table reports estimation results from a regression of of worker productivity index on EA-PGI of interacted with education level. All regressions control for first ten genetic PCs, gender, year of birth, calendar year and biobank indicators. Column (1) reports the baseline results. Column (2) additionally controls for detailed field of education (3-digit) indicators. Column (3) adds interaction between level and detailed field of education. Column (4) adds indicators for academic institution identifiers. Standard errors reported in parentheses.

	Dependent variable: Worker productivity index			
	(1)	(2)	(3)	(4)
Secondary education × EA-PGI	0.017*	0.010	0.011	0.007
	(0.007)	(0.007)	(0.007)	(0.007)
Tertiary education × EA-PGI	0.115***	0.086***	0.082***	0.016*
	(0.007)	(0.007)	(0.007)	(0.006)
Level	Yes	Yes	Yes	Yes
Field	No	Yes	Yes	Yes
Level × Field	No	No	Yes	No
Institution ID	No	No	No	Yes
Obs.	31,866	31,709	31,702	31,574
Avg. obs. per cell	15,933	352	145	23
Adj R2	0.198	0.251	0.256	0.330
RMSE	0.836	0.808	0.806	0.765

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.7: Years of education and EA-PGI, unconditional and conditional on parental EA-PGI among parent-offspring trios
The table reports estimation results from regression of predicted years of education (given highest qualification) on EA-PGI, unconditional and conditional on parental EA-PGI in parent-offspring trios. The top panel reports baseline estimations without controlling for parental EA-PGI. The bottom panel reports estimates conditional on parental EA-PGI. Column (1) uses full sample of parent-offspring trios ($N = 12{,}871$), including the cases where parental genotypes were imputed from parent-offspring duos and siblings. Column (2) uses subset of directly genotyped parent-offspring trios ($N = 4{,}586$). All estimations control for first ten genetic principal components, gender, year of birth, and biobank indicators. Standard errors reported in parentheses.

	Dependent var: predicted years of education	
	All family trios (1)	Directly genotyped trios (2)
Baseline without parental EA-PGI		
Own EA-PGI	0.553***	0.570***
	(0.016)	(0.027)
Constant	14.691***	13.741***
	(0.491)	(1.058)
Obs.	12 871	4 586
Controlling for parental EA-PGI		
Own EA-PGI	0.413***	0.441***
	(0.026)	(0.040)
Mother EA-PGI	0.128***	0.110***
	(0.021)	(0.030)
Father EA-PGI	0.093***	0.095***
	(0.021)	(0.030)
Constant	14.641***	13.717***
	(0.482)	(1.028)
Obs.	12 871	4 586

Table A.8: Cumulated lifetime income, own and parental EA-PGI, regression results
The table reports estimation results from regression of cumulated lifetime income on own and parental EA-PGI. Columns (1)-(3) report the baseline estimates without controlling for parental EA-PGI. Columns (4)-(6) report estimates conditional on parental EA-PGI. Columns (1) and (4) use full sample of parent-offspring trios (N = 12,871), columns (2) and (5) - subset of trios with secondary-educated offspring (N = 5,063) and columns (3) and (6) - subset of trios with tertiary-educated offspring (N = 7,808). All estimates control for first ten genetic principal components, gender, birth year, calendar year, and biobank indicators. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Standard errors reported in parentheses.

	Dependent variable: Cumulated income					
	Baseline			Controlling for parental EA-PGI		
	Pooled (1)	Secondary (2)	Tertiary (3)	Pooled (4)	Secondary (5)	Tertiary (6)
Own EA-PGI	13 724	-7 721	15 602	7 398	-3 613	5 048
	(1 804)	(2 132)	(2 546)	(2 746)	(3 478)	(3 833)
Mother EA-PGI				1 769	-5 002	4 658
				(2 163)	(2 680)	(3 016)
Father EA-PGI				8 211	-1 702	12 113
				(2 375)	(2 795)	(3 370)
Constant	414 682	270 156	507 419	411 300	270 147	501 314
	(61 295)	(58 545)	(84 159)	(61 302)	(57 824)	(84 234)
Obs.	12 871	5 063	7 808	12 871	5 063	7 808

Table A.9: Cumulated lifetime income by own and parental EA-PGI levels
The table reports adjusted lifetime income (up to 25 years since graduation) by EA-PGI percentiles, conditional on parental EA-PGI. Columns (1)-(3) use full sample of parent-offspring trios with tertiary-educated offspring ($N = 7{,}808$). Columns (4)-(6) use subset of directly genotyped parent-offspring trios with tertiary-educated offspring ($N = 2{,}704$). Columns (1) and (4) report cumulated lifetime income at 10th and 90th percentiles of own EA-PGI, holding maternal and paternal EA-PGI constant. Columns (2) and (5) report cumulated lifetime income at 10th and 90th percentiles of maternal EA-PGI, holding own and paternal EA-PGI constant. Columns (3) and (6) report cumulated lifetime income at 10th and 90th percentiles of paternal EA-PGI, holding own and maternal EA-PGI constant. Average lifetime income adjusted by regressing cumulated income on the set of own, maternal and paternal EA-PGI and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Standard errors reported in parentheses.

	Dependent variable: Cumulated income					
	All family trios			Directly genotyped trios		
	Own EA-PGI (1)	Mother EA-PGI (2)	Father EA-PGI (3)	Own EA-PGI (4)	Mother EA-PGI (5)	Father EA-PGI (6)
10th percentile	339 181	339 879	330 699	350 648	342 498	336 264
	(5 690)	(4 311)	(4 459)	(9 782)	(6 842)	(6 792)
90th percentile	352 172	351 466	360 946	350 086	358 149	364 690
	(5 198)	(4 500)	(5 126)	(8 547)	(6 833)	(7 444)
90-10 gap	12 991	11 587	30 247	-562	15 650	28 426
	(9 866)	(7 502)	(8 414)	(16 426)	(10 933)	(11 639)
Obs.	7 808	7 808	7 808	2 704	2 704	2 704

Table A.10: Cumulated lifetime income of secondary-educated individuals by EA-PGI level, conditional on parental EA-PGI among parent-offspring trios
The table reports adjusted lifetime income (up to 25 years since graduation) by EA-PGI percentiles. The estimation sample is restricted to parent-offspring trios with secondary-educated offspring ($N = 5{,}063$). Column (1) reports the estimates in the baseline specification without controlling for parental EA-PGI. Columns (2)-(4) report average cumulated lifetime income by own, maternal and paternal EA-PGI percentiles, respectively, conditional on parental EA-PGI. Average lifetime income adjusted by regressing cumulated income on EA-PGI and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Standard errors reported in parentheses.

	Dependent variable: Cumulated income			
	Baseline	Controlling for parental EA-PGI		
	Own EA-PGI	Own EA-PGI	Mother EA-PGI	Father EA-PGI
	(1)	(2)	(3)	(4)
EA-PGI percentiles				
10th	265 492	259 961	261 546	257 400
	(2 782)	(4 621)	(3 938)	(3 954)
20th	262 054	258 352	259 384	256 677
	(2 229)	(3 323)	(3 036)	(3 029)
30th	259 582	257 195	257 866	256 156
	(2 022)	(2 568)	(2 532)	(2 503)
40th	257 496	256 219	256 508	255 710
	(2 017)	(2 184)	(2 240)	(2 223)
50th	255 614	255 338	255 320	255 311
	(2 150)	(2 154)	(2 161)	(2 159)
60th	253 639	254 413	254 078	254 906
	(2 405)	(2 459)	(2 274)	(2 294)
70th	251 508	253 416	252 817	254 471
	(2 775)	(3 057)	(2 566)	(2 624)
80th	248 921	252 206	251 177	253 929
	(3 311)	(3 978)	(3 130)	(3 216)
90th	245 621	250 661	249 103	253 149
	(4 074)	(5 294)	(4 008)	(4 253)
Obs.	5 063	5 063	5 063	5 063

Table A.11: Cumulated lifetime income by own EA-PGI level, education, genotype sample and weighting scheme

The table reports adjusted lifetime income (up to 25 years since graduation) by own EA-PGI percentiles. Panel A uses full analysis sample (N = 51,056 in pooled, N = 18,692 in secondary-educated and N = 32,364 in tertiary-educated subsamples), while Panel B uses sample of parent-offspring trios (N = in pooled, N = 5,063 in secondary-educated and N = 7,808 in tertiary-educated subsamples). Columns (1), (3) and (5) report unweighted estimates. Columns (2), (4) and (6) report weighted estimates using inverse probability weights. Average lifetime income adjusted by regressing cumulated income on EA-PGI and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Panel B additionally controls for maternal and paternal EA-PGI. Income discounted to obtain its present value upon graduation (see Section 1.2 for additional information). Inverse probability weights are estimated to balance the sample according to year of birth and graduation year (fully interacted with highest education level and gender), and rural area indicator fully interacted with gender. Standard errors reported in parentheses.

	Dependent variable: Cumulated income					
	Pooled		Secondary		Tertiary	
	Unweighted (1)	Weighted (2)	Unweighted (3)	Weighted (4)	Unweighted (5)	Weighted (6)
Panel A: Genotyped sample						
10th percentile	309 659	291 728	262 386	257 996	346 194	331 362
	(1 306)	(1 286)	(1 429)	(1 501)	(1 944)	(1 920)
50th percentile	329 893	308 756	255 422	249 549	368 728	350 947
	(857)	(832)	(1 116)	(1 157)	(1 137)	(1 105)
90th percentile	350 418	325 930	248 358	241 029	391 585	370 700
	(1 591)	(1 525)	(2 120)	(2 185)	(2 006)	(1 938)
Obs.	51 056	51 056	18 692	18 692	32 364	32 364
Panel B: Family trio sample (conditional on parental EA-PGI)						
10th percentile	303 725	304 546	259 961	263 787	339 181	345 745
	(3 906)	(4 417)	(4 621)	(5 261)	(5 690)	(6 559)
50th percentile	313 190	313 886	255 338	258 422	345 639	351 212
	(1 697)	(1 981)	(2 154)	(2 388)	(2 320)	(2 755)
90th percentile	322 764	323 347	250 661	252 986	352 172	356 750
	(3 934)	(4 502)	(5 294)	(5 717)	(5 198)	(6 114)
Obs.	12 871	12 871	5 063	5 063	7 808	7 808

Table A.12: Predictive power of EA-PGI for years of education
The table reports estimation results from regression of predicted years of education (given highest qualification) on EA-PGI. Column (1) reports estimates without EA-PGI, and column (2) - with EA-PGI. All estimations controls for gender fully interacted with year of birth indicators, biobank indicator and first ten genetic principal components. Standard errors reported in parentheses.

	Dependent variable: predicted years of education	
	(1)	(2)
EA-PGI		0.530***
		(0.008)
Constant	14.984***	14.979***
	(0.202)	(0.194)
Obs.	51,056	51,056
R^2	0.086	0.158
Adj. R^2	0.085	0.157
Incremental R^2		0.071

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$