Genetics, Educational Attainment, and Socioeconomic Trajectories over the Life Cycle*

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 51,735 graduates (972,897 person-year observations) followed annually since graduation for up to 25 years, we report four findings. First, higher EA-PGI does not predict higher income at labor market entry. Instead, it strongly predicts subsequent income growth, but only among higher-educated people: tertiary-educated graduates at the 90th percentile earn €45 612 (13.2 %) higher lifetime income than those at the 10th percentile. Second, EA-PGI does not predict 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 reduces the lifetime income gap to €13 003. Finally, the above results are unlikely to be mediated by health, since EA-PGI only weakly predicts disease burden.

*We would like to thank Adrian Adermon, Olof Åslund, Pietro Biroli, Cristina Bratu, Lena Hensvik, Paul Hufe, Andrea Ichino, Francis Kramarz, Martin Nybom, Oskar Nordström Skans, Gerard van den Berg, and participants to the European Social Science Genetics Network (ESSGN) 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 are grateful to Blood Service Biobank and to THL Biobank for issuing the data permits to utilize the genetic data. Corresponding authors: Andrea Ganna (andrea.ganna@helsinki.fi), Stefano Lombardi (stefano.lombardi@vatt.fi).

1 Introduction

Understanding the origins, timing 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 (Cunha and Heckman, 2007). For a child being born into an affluent family, having a low genetic predisposition to disease, or living in an economically striving area is a matter of chance. However, such factors shape socioeconomic status and, consequently, can affect equality of opportunity both within and across generations (Roemer and Trannoy, 2016).

Economists have long sought to identify the drivers socioeconomic inequality, traditionally focusing on income disparities and, more recently, extending to differences in health (Black et al., 2024). Income and health outcomes arise from a complex interplay between genetics and environmental factors over the life cycle. The idea that nature and nurture jointly determine outcomes is central to economic models of skill formation (Cunha and Heckman, 2007), in which genetic endowments are considered as propensities that occur via environmental interactions.[1] 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, including long-term outcomes such as wealth at retirement, exhibit substantial heritability (Barth, Papageorge, and Thom, 2020; Harden and Koellinger, 2020; Rustichini et al., 2023; Carvalho, 2025). Educational attainment (EA) has been a primary focus of this growing literature, facilitated by large-scale genome-wide association studies (GWAS) and validated PGI (Okbay et al., 2022). These scores capture both direct genetic effects – randomly inherited at conception – and indirect effects, such as environmentally mediated influences of parental genetics, i.e. genetic nurture (Kong et al., 2018).

Despite these advances, research provides limited insight into how genetic influences on socioeconomic status evolve over the life course. Most studies rely on cross-sectional data, providing static estimates that are silent about the temporal dynamics of inequality.[2] It also remains unclear whether, and to what extent, genetic effects are driven by certain population subgroups, such as by low- or high-educated individuals. Finally, despite a large literature in economics points towards the role of employers in determining wage inequality (see Kline, 2024, for a review), we do not know the role of firms in mediating any existing income gaps by genetics.

[1] For reviews on gene-environment interactions, see Biroli et al. (2025) and Pereira et al. (2022).

[2] One recent exception is Akimova et al. (2025), who analyze career trajectories in the UK biobank.

Understanding these mediating mechanisms is crucial for identifying when and how environmental interventions may amplify or mitigate genetic predispositions. Prior evidence indicates that genetic effects on SES vary across contexts, for example, before and after the collapse of communism in Estonia (Rimfeld et al., 2018) and by family background, with children of lower genetic endowments achieving better in high-SES families, consistent with a compensatory role of family resources (Ghirardi et al., 2024).

To address this gap, we combine comprehensive Finnish administrative register data on education, income, demographics, and health with polygenic indices for educational attainment (EA-PGI). Each fresh graduate is repeatedly observed throughout their prime working age since graduation until up to 25 years later, and workers are matched each year to their current employer. With this matched employer-employee panel data at hand, we adopt a trajectory-based approach to examine how the genetic predisposition to education, captured by the EA-PGI, influences the evolution of labor income, employer quality, and health over the life cycle, and whether the genetics gradients in these outcomes vary across educational groups.

Overall, this paper analyzes the mechanisms underlying differences in socioeconomic trajectories by analyzing the role of parental PGIs, by assessing the contribution of employers and labor market dynamics to the development of inequality, and by examining whether health functions as an intermediate factor correlated with educational attainment, and thus potentially narrowing inequality gaps.

We advance the literature in the following important ways. First, we employ a substantially larger sample with genetic information (including on parents) than previous studies. This, in combination with decades of register data, allows us to generate highly precise estimates on the drivers of socioeconomic disparities. Second, we consider income as a primary measure of socioeconomic inequality and exploit unusually rich information from tax records on labor income on the continuum. This source of information, which is still extremely rare in the sociogenomics literature, allows us to avoid selective self-response, missing not-at-random, and measurement error that could otherwise severely bias our estimates. Third, having access to repeated measurements of both workers, employers, and their link over time, opens the way to leveraging panel data methods to account for firm- and worker-level 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.

2 Results

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

The study includes 51 736 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, capital, and benefits, and educational records (education level and field and school/university identifiers). This information spans thirty years (1987–2019). We apply a dynamic modelling framework 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. To enhance the transparency of the results, all estimates are presented in both graphical and tabular form.

Comparing earned income trajectories between the 10th and 90th percentile of the EA-PGI distribution over 25 years since graduation, Figure 1a shows that income levels are initially very similar across groups. However, trajectories begin to diverge substantially over time. The gap in average income between 90th and 10th percentile of EA-PGI 10 years from graduation is €3278 and widens further to €7571 by 25 years from graduation (the average yearly earnings in the sample 10 and 25 years after graduation is €24 932 and €34 858, respectively). The cumulated income over the 25 years after graduation is €309 438 for individuals at the 10th percentile and €350 395 for those at the 90th percentile of the EA-PGI distribution.

Strikingly, this genetic gradient in income varies markedly by educational attainment. Figures 1b and 1c show that the income gap by genetics is driven entirely by individuals with tertiary education, 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.

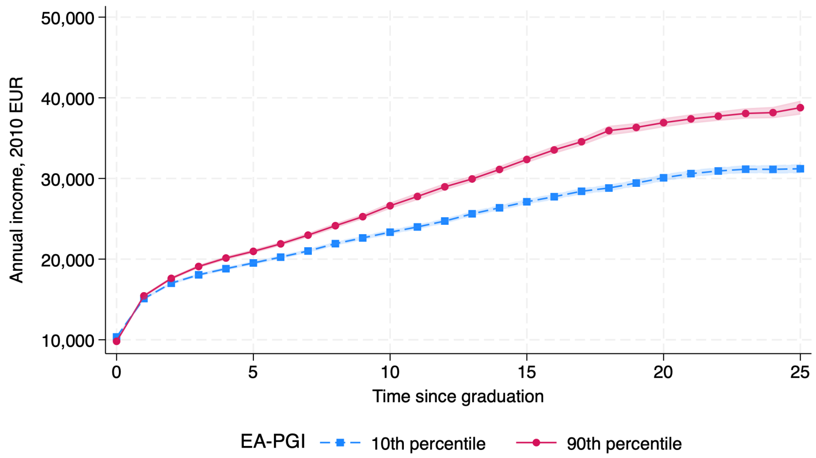
Table 1 reports the cumulated income over the 25 years following graduation by EA-PGI percentiles and educational attainment. In the pooled sample (column 1), the gap between the 10th and 90th percentile is €40 957, or approximately 13.2 %. Column 3 of the table confirms that the income gap is entirely driven by the tertiary-educated individuals.

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

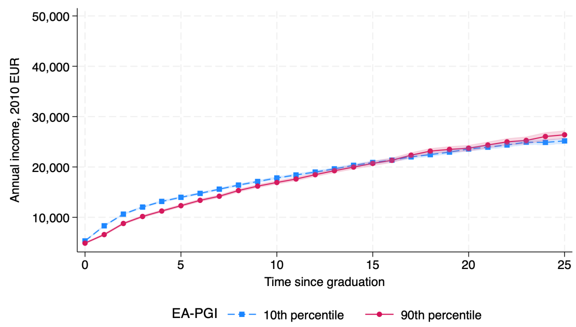
To analyze the extent to which the income gradient in EA-PGI reflects higher productivity in the labor market, we estimate an Abowd–Kramarz–Margolis (AKM), the

Figure 1: Average annual income by EA-PGI level, over time and by education

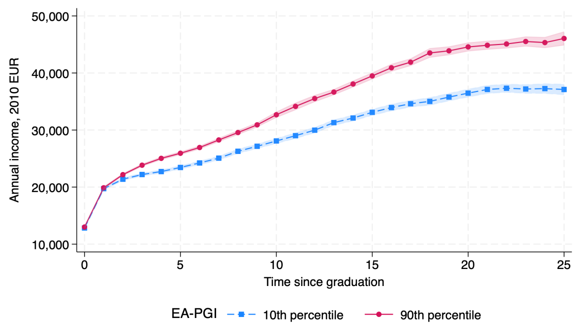
(a) Pooled



(b) Secondary



(c) Tertiary



Notes: Adjusted average annual income (in 2010 EUR) over time. Panel A uses full analysis sample, while Panels B and C use subset of workers based on their highest qualification being either secondary or tertiary degree, respectively. The lines correspond to 10th and 90th percentiles 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.

Table 1: Cumulative lifetime income by EA-PGI percentiles

	Dependent variable: cumulated income		
	Pooled	Secondary	Tertiary
EA-PGI percentiles			
10th	309 438	262 462	345 947
	(1 316)	(1 440)	(1 959)
20th	316 387	260 071	353 686
	(1 040)	(1 156)	(1 543)
30th	321 424	258 337	359 296
	(900)	(1 046)	(1 307)
40th	325 748	256 849	364 111
	(844)	(1 041)	(1 176)
50th	329 829	255 444	368 656
	(856)	(1 114)	(1 137)
60th	333 865	254 055	373 151
	(929)	(1 247)	(1 187)
70th	338 193	252 566	377 971
	(1 060)	(1 438)	(1 327)
80th	343 367	250 785	383 733
	(1 264)	(1 708)	(1 578)
90th	350 395	248 366	391 560
	(1 590)	(2 119)	(2 004)
Obs.	51 056	18 692	32 364

Notes: The table reports adjusted lifetime income (up to 25 years since graduation) by EA-PGI percentiles. 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 net present value upon graduation (see Section 4.2 for additional information). Standard errors reported in parentheses.

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 fixed effects. After model estimation, each worker is assigned a corresponding estimated fixed effect, which we refer to as *worker productivity index*, as it captures 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.[3]

Figure 2 shows 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 only for the tertiary-educated workers. For this group, a one standard deviation increase in EA-PGI is associated with a 0.166 standard deviation higher worker productivity, compared to a 0.022 increase among the secondary-educated individuals. Hence, the high-EA-PGI people 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 analyze whether the difference in slope coefficients between the two education groups persists when adjusting for a gradually richer set of controls (education track, field, track-by-field, and school or university fixed effects). Although the association between EA-PGI and worker productivity attenuates (Appendix Table A.6), among the tertiary-educated workers it remains statistically significant, even after adjusting for this extensive set of controls.

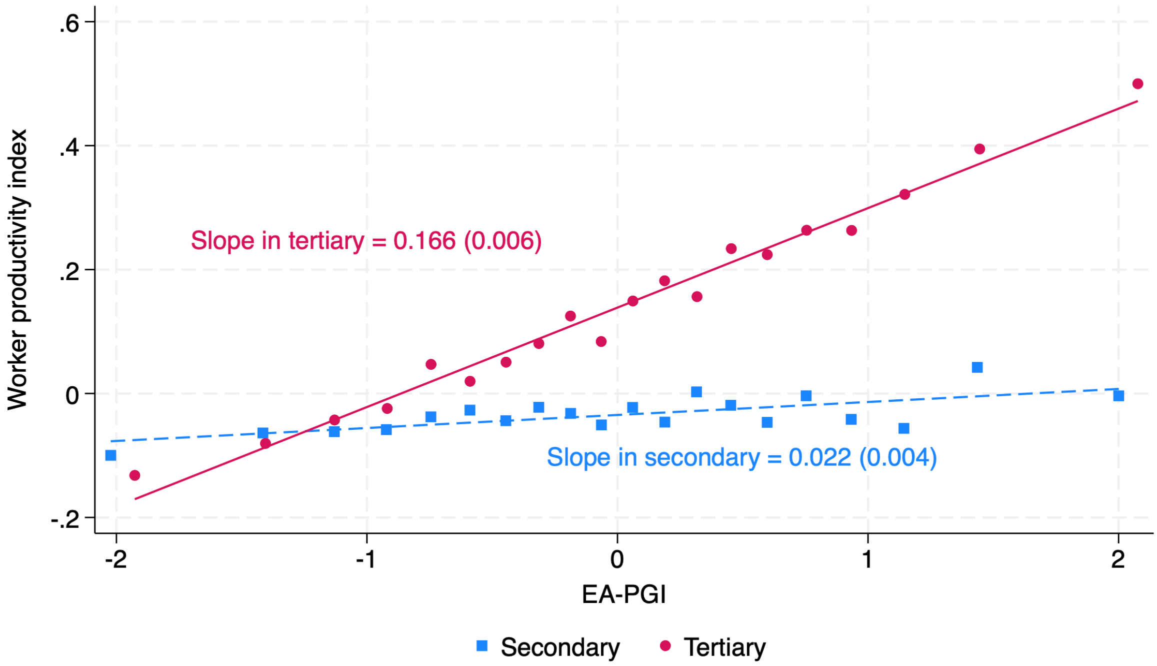
2.3 High EA-PGI individuals transition more rapidly and more frequently to higher-quality firms.

We next investigate the role of firms in shaping income differences across levels of EA-PGI among tertiary-educated individuals. Figure 3a shows that workers with higher EA-PGI change employers slightly more frequently. To assess the quality dimension of these moves, we use the AKM firm fixed effects (previously estimated along with the workers' fixed effects) standardized to have mean 0 and standard deviation 1. We consider them as a *firm quality index*, since for each employer they capture 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.

Figure 3b shows that, on average, higher EA-PGI individuals transition to significantly higher-quality firms. Interestingly, the quality of their first employer is not

[3] The model controls for non-linear age and calendar time fully-interacted with education. Estimation details are provided in Section 4.2 and descriptive statistics in Appendix Tables A.3 and A.4.

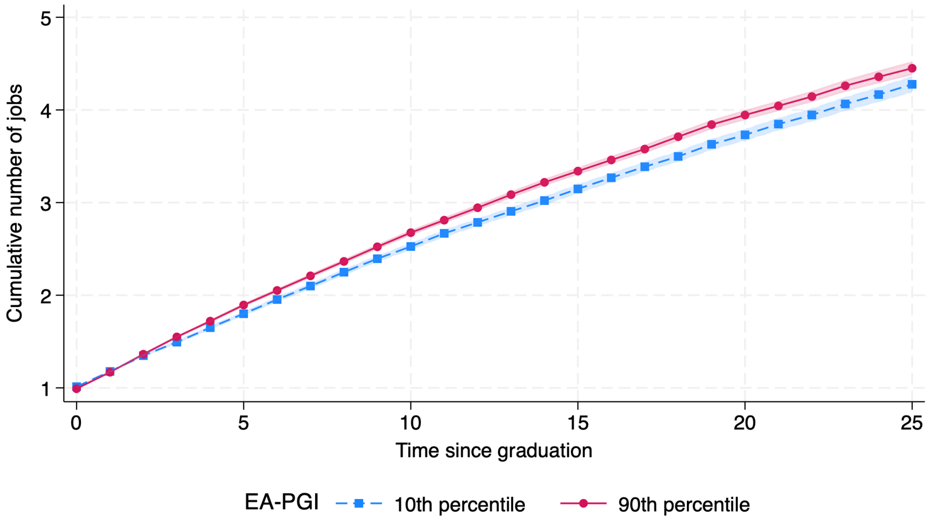
Figure 2: Relationship between worker productivity and EA-PGI, by education group



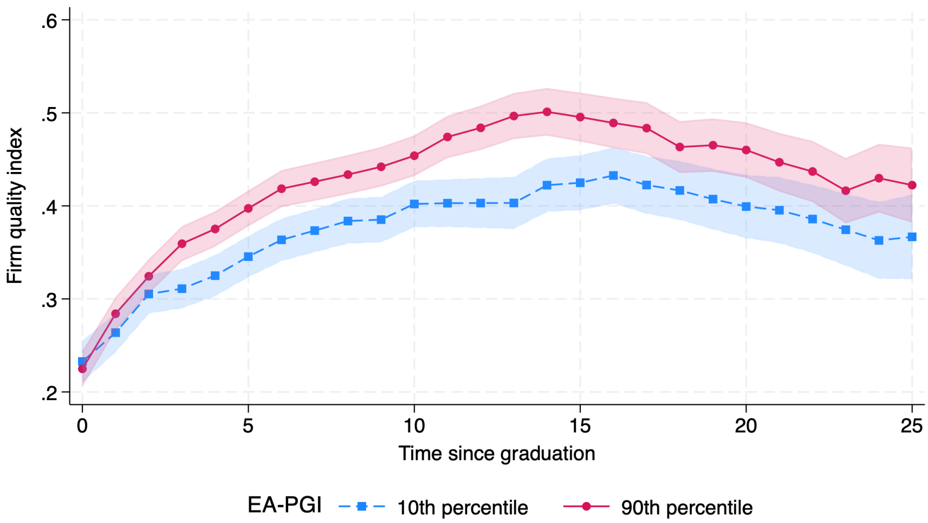
Note: The vertical axis reports a worker productivity measure estimated via an AKM model (details reported in Section 4.2); the horizontal axis reports ventiles of the EA-PGI distribution. The scatterplot shows the relation between these two quantities standardized to have mean 0 and standard deviation 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 55 435 individuals. 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 indicators and first ten genetic PCs.

Figure 3: Employer quality and number of jobs since labor market entry

(a) Number of jobs since graduation



(b) Firm quality since graduation



Notes: Panel A plots the average number of employment spells over time since graduation at 10th and 90th percentiles of EA-PGI. Panel B plots the average firm quality index over time since graduation at 10th and 90th percentiles of EA-PGI. Both 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. The shaded areas correspond to 95% CIs.

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.

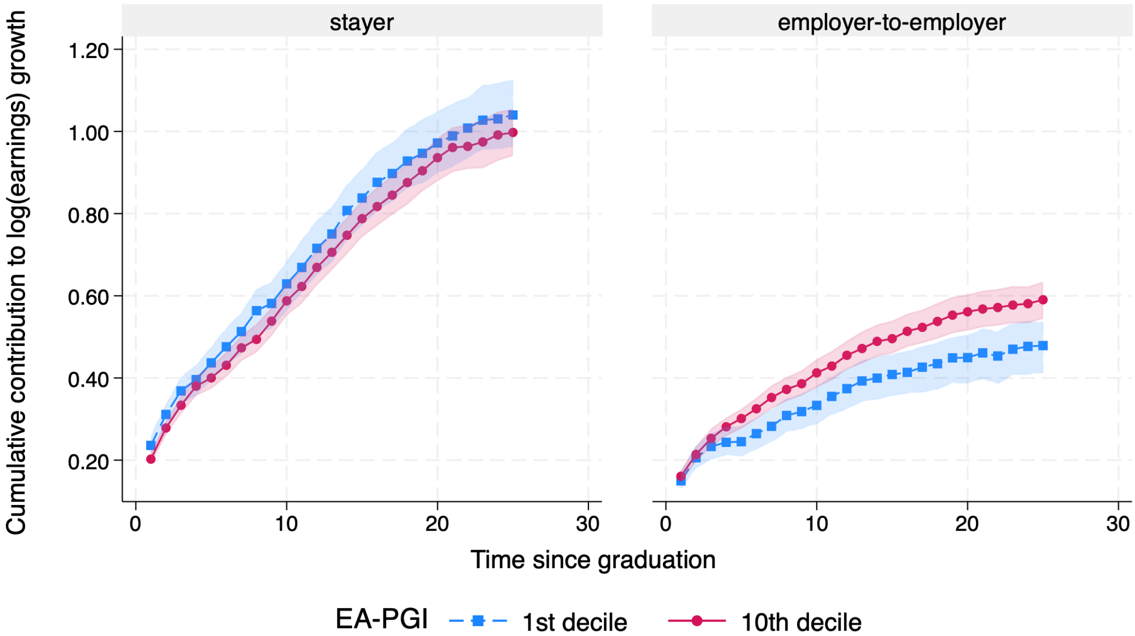
In contrast, among individuals with secondary education, firm quality trajectories appear similar across EA-PGI levels (Appendix Figure A.5). Moreover, individuals in the bottom decile of the EA-PGI distribution, regardless of whether they hold tertiary or secondary education, exhibit strikingly similar mobility patterns with respect to firm quality. These findings suggest 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.

2.4 Income growth disparities between high and low EA-PGI individuals are attributable to differences in mobility across jobs.

To better understand the drivers of the income differences between EA-PGI groups, we follow Hahn, Hyatt, and Janicki (2021) and decompose labor income changes between any two consecutive years since graduation into within-firm increases (stayers), employer-to-employer mobility, and transitions into or out of unemployment. Figure 4 shows the contribution of the stayers and job-to-job movers components to the cumulative log-income growth over time and by EA-PGI group, confirming that, on average, both are associated with earnings gains over the life cycle. Details on the decomposition are reported in Section 4.2, while Figure A.6 in Appendix shows all four components (including transitions in and out of unemployment) and the corresponding workers shares over time.

The right panel of Figure 4 shows that the contribution of between-firm mobility to earnings growth becomes relatively more important over time for people at the 90th percentile of EA-PGI compared to those at the 10th percentile. When comparing the two EA-PGI groups, the gap in the job-to-job contribution to earnings growth cumulated over 25 years is of about 0.1 log-points (approximately 10%). This suggests that a substantial part of income divergence is driven by firm-to-firm mobility. Transitions into and out of unemployment account for only a negligible share of the earnings growth disparities across EA-PGI groups (Figure A.6).

Figure 4: Decomposition of total cumulative growth in log(earnings)



Notes: the figure reports the results of the decomposition of total earnings growth to within-firm growth, employer-to-employer switches, and mobility to/from non-employment. The figure shows the cumulative contributions of within-firm earnings growth and earnings growth attributed to employer-to-employer switches over time since graduation at first and tenth EA-PGI decile. The full decomposition including growth attributed to exit to and entrance from non-employment is presented in Appendix Figure A.6a. The decomposition is applied to log annual earnings and follows Hahn, Hyatt, and Janicki (2021). The sample for the decomposition is restricted to tertiary-educated workers. 95% Confidence Intervals 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.

2.5 Parental EA-PGI, in particular that of fathers, predicts the income trajectories of the 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 10th percentile of EA-PGI show remarkably lower socioeconomic status (parental education) than those at the 90th percentile (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 12 918 parent–offspring trios. Of these, 4587 were directly genotyped, while the remainder were imputed based on 10 295 duos and 25 514 sibling pairs.[4]

Figure 5 presents the main results separately for secondary-educated (left panel) and tertiary-educated individuals (right panel), comparing estimates with and without parental EA-PGI controls. Among secondary-educated individuals, the income gap across EA-PGI percentiles remains negligible, consistent with our earlier conclusions. On the other hand, for the tertiary-educated individuals the parental EA-PGI accounts for a meaningful share of the observed gap between the 90th and 10th percentiles of the offspring EA-PGI (Figures 5b vs. 5d). Following the approach outlined in Table 1, we estimate that controlling for parental EA-PGI reduces this gap by approximately 71 % (see Appendix Table 2).[5] Note that, as expected (and explicitly modelled and analyzed in Rustichini et al., 2023), controlling for parental PGI-EA captures outcome variation due to parental genetics, as well as family background characteristics unevenly distributed across the EA-PGI groups (Table A.5).

2.6 EA-PGI is weakly associated with disease burden trajectories

Health is a well-documented determinant of education and income disparities (e.g., Pallesen et al., 2024; Newman, Gordon, and Mendes, 2025), which makes it a possible intermediate channel between EA-PGI and income. To examine whether the correlation between EA-PGI and income is mediated by disease burden, we compute the Charlson Comorbidity Index (CCI), which records the first occurrence of 17 major chronic conditions over the life cycle (Charlson et al., 1987; Deyo, Cherkin, and Ciol, 1992). We then use it as an outcome in our trajectories model.

Figure 6 shows that the CCI is, on average, lower among individuals with tertiary compared to secondary education, reflecting a well-established lower incidence of ma-

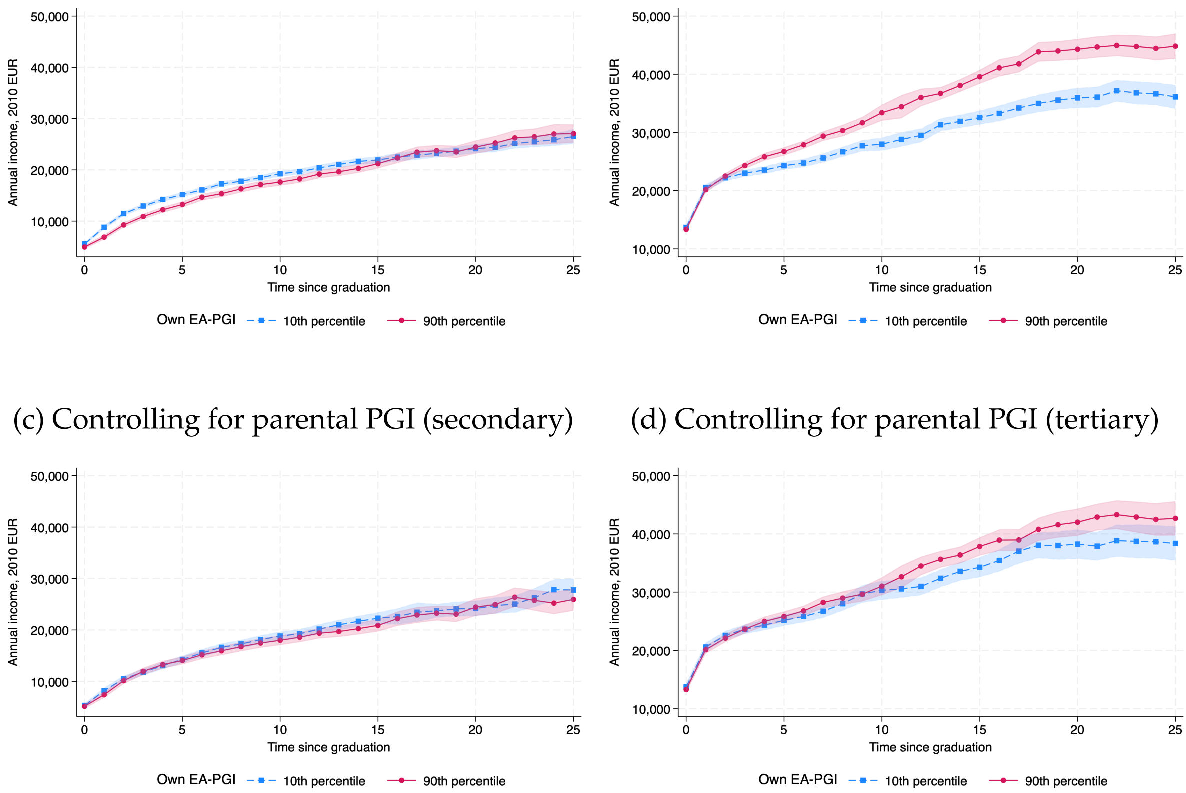
[4] We show that EA-PGI strongly predicts years of education. Consistent with previous findings (e.g., Wang et al., 2021), the effect of offspring EA-PGI on years of education is attenuated but remains significant once paternal and maternal EA-PGIs are included as controls; see Appendix Table A.7.

[5] Appendix Figure A.8 further indicates that paternal EA-PGI contributes more strongly to the income difference between the 90th and 10th EA-PGI percentiles than maternal EA-PGI.

Figure 5: Average annual income by EA-PGI level, over time and by education, unconditional an conditional on parental PGI

(a) Baseline without parental PGI (secondary) (b) Baseline without parental PGI (tertiary)

(c) Controlling for parental PGI (secondary) (d) Controlling for parental PGI (tertiary)



Notes: the figure plots average income trajectories over time since graduation at 10th and 90th percentiles of EA-PGI. Panels A and B plot baseline trajectories among secondary- and tertiary-educated workers 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 trajectories are estimated from a regression of annual earnings 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 12 918 parent-offspring trios (directly genotyped and imputed). The shaded areas correspond to 95% CIs.

Table 2: Cumulative lifetime income by EA-PGI percentiles, controlling for parental EA-PGI

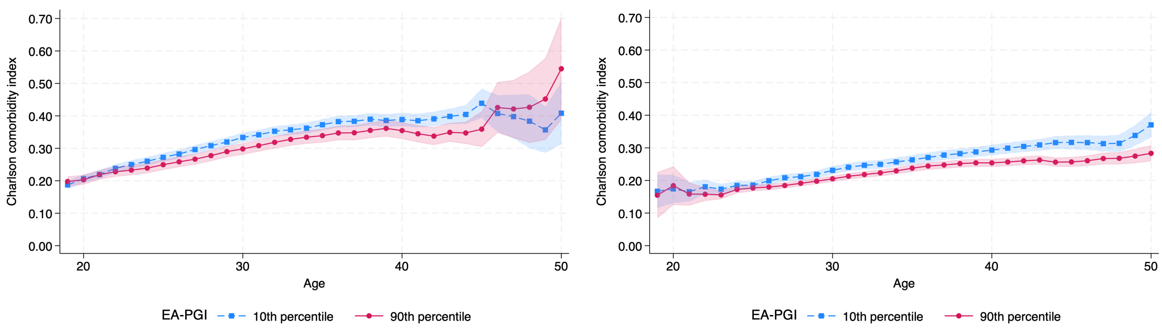
	Dependent variable: cumulated income					
	Baseline			Controlling for parental PGI		
	Pooled	Secondary	Tertiary	Pooled	Secondary	Tertiary
EA-PGI percentiles						
10th	295 529	265 504	325 772	303 711	259 965	339 170
	(2 574)	(2 785)	(4 018)	(3 910)	(4 625)	(5 695)
20th	301 655	262 058	332 736	307 013	258 352	341 423
	(2 026)	(2 229)	(3 161)	(2 856)	(3 324)	(4 187)
30th	306 025	259 599	337 704	309 369	257 202	343 030
	(1 760)	(2 023)	(2 680)	(2 214)	(2 572)	(3 238)
40th	309 751	257 502	341 940	311 378	256 221	344 401
	(1 661)	(2 017)	(2 413)	(1 827)	(2 184)	(2 612)
50th	313 091	255 624	345 736	313 178	255 342	345 629
	(1 693)	(2 149)	(2 328)	(1 697)	(2 153)	(2 321)
60th	316 614	253 641	349 742	315 077	254 414	346 925
	(1 842)	(2 404)	(2 413)	(1 834)	(2 459)	(2 399)
70th	320 400	251 511	354 046	317 118	253 417	348 318
	(2 107)	(2 775)	(2 685)	(2 235)	(3 056)	(2 883)
80th	325 022	248 911	359 300	319 610	252 200	350 018
	(2 528)	(3 313)	(3 200)	(2 922)	(3 982)	(3 803)
90th	330 882	245 615	365 961	322 768	250 658	352 173
	(3 150)	(4 076)	(4 024)	(3 936)	(5 297)	(5 201)
Obs.	12 871	5 063	7 808	12 871	5 063	7 808

Notes: The table reports adjusted lifetime income (up to 25 years since graduation) by EA-PGI percentiles. Average lifetime income adjusted by regressing cumulated income on EA-PGI controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The first three columns replicate the baseline estimation without including parental EA-PGI in the subsample of family trios. The second three columns control for both parents' EA-PGI. Income discounted to obtain its net present value upon graduation (see Section 4.2 for additional information). Standard errors reported in parentheses.

jor chronic diseases among high-educated individuals (Agardh et al., 2011; Tillmann et al., 2017; Vaccarella et al., 2023). Consistent with this pattern, Figure 6 further indicates that higher EA-PGI is significantly associated with a lower cumulative disease burden as measured by the CCI. Importantly, the extent of this association between EA-PGI and CCI is very similar across both education groups. This suggests that the association between EA-PGI and income observed only among tertiary-educated individuals, is not directly explained by increased differences in disease burden.

Figure 6: Average health indices by EA-PGI levels, over time and by education

(a) Secondary (b) Tertiary



Notes: the figure plots average Charlson Comorbidity Index over time since graduation at 10th and 90th percentiles of EA-PGI. Panels A and B report the results for secondary- and tertiary-educated workers, respectively. The trajectories are estimated from a regression of the 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.

3 Discussion

In the first part of the analysis we focus on the labor income component of socioeconomic status, following individuals annually from the year of graduation through 1987–2019. The analysis of labor income since graduation is a central contribution of this paper, motivated by the fact that, while interpersonal differences can be determined as early as in utero (Almond and Currie, 2011), graduation constitutes a pivotal stage at which inequalities begin to materialize in life (von Wachter, 2020).

Our focus on labor income is motivated by the fact that it is a well-defined, downstream measure of socioeconomic status that reflects differences in genetic endowments, educational attainment, health, and skills. As opposed to inheritances, labor income (along with government transfers) accounts for the lion's share of the monetary inflows that people accumulate during lifetime, making it a fundamental component of wealth (Black et al., 2025). In addition, income inequality is a central and active research area in economics (see Kline, 2024 for a review), allowing us to build on well-established methodologies while incorporating genetic endowments into the same analytical framework. Finally, income is strongly correlated with health, which is the other primary component of socioeconomic status and additional outcome in our analysis.

Our analysis starts by examining how genetic endowments predictive of educational attainment shape income trajectories over the life course. A first notable result is that workers earn similar income upon labor market entry, irrespective of their EA-PGI level. After labor market entry, and only for the tertiary-educated people, the income differences between workers at the 90th and 10th percentile of EA-PGI keep widening up over time, resulting in an income gap cumulated over the 25 years since graduation of 13.2 %. The similar income levels upon entry followed by the observed divergence is consistent with an initial lack of worker's sorting according to ability, followed by relatively quick employer learning about worker's productivity (Farber and Gibbons, 1996).

The fact that EA-PGI explains only the income trajectories of the tertiary-educated people reflects a combination of two forces. People with relatively high EA-PGI experience higher economic returns to ability in certain sectors and occupations that are accessed via tertiary education degrees. At the same time, they sort into different education tracks and levels based on EA-PGI itself. 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.

As a starting point, 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 association persists even after controlling for detailed education-related mediating factors (education field, track, and institution id). The fact that EA-PGI relates to labor market productivity only for the higher-educated people raises the question of whether some of the high-ability people with secondary education should be encouraged (or economically supported) to continue into tertiary education. This project does not allow us to answer this question, but we note that the answer should necessarily take into account the socioeconomic background of pupils (see Ichino, Rustichini, and Zanella, 2024). We return to the importance of family background later.

Our analysis also shows that 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. Despite the absence of initial sorting, genetic endowments subsequently matter for transitions to higher-paying employers, indicating that employer learning about workers' productivity and job mobility are important channels underlying the genetic gradient in income. Consistent with this, we observe significantly steeper firm-quality trajectories among workers at the 90th EA-PGI percentile, whereas those at the 10th percentile exhibit flat employer-quality profiles. While these results partly reflect sorting into better occupations, occupational choice alone is unlikely to capture the income gains driven by the (much more granular) differences in *employer* quality that we show. This is because of the typically large variation of firm quality within occupations (e.g., Card, Heining, and Kline, 2013).

By Mendel's laws, parents of children with higher EA-PGI must have higher EA-PGI themselves, which results in higher parental education. This is confirmed in our sample, where people at the 10th percentile of EA-PGI show remarkably lower socioeconomic status than those at the 90th percentile (Table A.5), a pattern that can be exacerbated by assortative mating.[6] We therefore examine the role of parental EA-PGIs in accounting for the observed income gap across offspring's EA-PGI levels. This approach isolates the direct genetic contribution to income by netting out indirect genetics and environmental channels. Controlling for parental EA-PGI reduces the income gap by approximately 71 %, indicating that indirect effects play an important role in shaping income development.

Interestingly, large part of the EA-PGI channel is explained by fathers' EA-PGI (rather than mothers'). This is consistent with maternal resources (such as active time spent with the child) being more relevant for early childhood development, whereas parental resources in the form of income (more often provided by fathers) tend to be more important at the later stages of development that we focus on in our analysis

[6] See Rustichini et al. (2023) for similar empirical patterns and for a theoretical framework of parental investments and intergenerational mobility that embeds a genetic analysis of skill transmission.

(e.g., Del Boca, Flinn, and Wiswall, 2014).

Finally, when we examine whether health constitutes an intermediate channel linking EA-PGI to income, we find that the association of a cumulative disease-burden index with EA-PGI is similar across education groups. This uggests that health is unlikely to mediate the strong EA-PGI–income relationship.

This paper contributes to a growing literature at the intersection of economics and human genetics that seeks to understand how nature and nurture jointly determine human differences and socioeconomic inequality. Existing cross-sectional studies in this field typically rely on outcomes that either lack income information altogether or use self-reported measures, which are subject to measurement error and bias. To the best of our knowledge, this is the first study to analyze income trajectories using register-based data that link individuals to employers over time through a matched employer-employee structure. The employer-employee link and repeated measurements over time allow us to decompose how firm- and worker-level factors account for interpersonal income differences due to genetics. In doing so, we builds a novel bridge with the modern economics literature on income inequality (Card et al., 2018; Song et al., 2018; Kline, 2024).

Taken together, our results suggest that genetic potential is most strongly expressed among tertiary-educated people – particularly those with an academic track degree – through labor market transitions towards higher-quality employers. These employers pay on average high wages, thereby contributing to the income inequality along genetics documented in the income analysis. Our results also show that indirect genetic effects and parental background explain a relevant part of the effect of EA-PGI on income.

4 Data and Methods

4.1 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 24 600 from the Blood Donor study.

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 %[7]

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

of the variance in years of education—slightly lower than estimates from previous studies Lee, Wedow, Okbay, et al., 2018, possibly due to differences in cohort composition or educational classification.

On average, our sample appears 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 population of labor market entrants (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.9).

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 annual total income and income by source (labor income, capital income, and income transfers and benefits). We include people with zero income in the income analysis, thereby avoiding conditioning on employment. All monetary values are deflated to 2010 EUR.

The registers further include demographic variables (gender, year of birth) and detailed educational information (highest degree, 2-digit field, vocational vs. academic track). Occupation codes (4-digit) are available annually from 2004 and industry codes from 1987. Both have been harmonized for consistency over time, using official lookup tables provided by the Statistics Finland.

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 used to complement the coverage. Both registers contain patient identifiers, care episode details, and one or more discharge diagnoses.

4.2 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 Appendix Figure A.1. To control for ancestry and population stratification, we compute the first 10 principal components of the genetic data and include them as covariates in all analyses that do not control for parental genetics.

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) model (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[8] for individual i in year t; $\mathbf{X}_{it}$ includes 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 ability (unobserved heterogeneity) and $\hat{\psi}_{J(i,t)}$ offers firm-specific wage premia, interpreted as firm quality. The estimated fixed effects are interpretable in relative terms: workers with higher fixed effects earn more across firms relative to others, 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 non parametrically. The estimation sample includes all full-time employees aged 20–60. For multiple employment spells, 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. Employment spells in

[8] 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), 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.

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 (Figure A.10).

We summarize worker and firm fixed effects (e.g., by percentiles) and link them to the genotyped sample. Appendix A 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 Trajectory Model

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

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

where y_{icmt} is total annual income of individual i, 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 gender, ten genetic principal components and biobank indicator (THL or Blood Donors). The coefficients of interest are β_t, which capture the income differential for EA-PGI over time since graduation either from secondary or tertiary education. Since PGIs are randomly assigned at conception, β_t has a causal interpretation, conditional on adequate control for population stratification via X_i. Residual stratification may remain, which 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.[9]

For graphical results and in our baseline analysis, we present average income trajectories evaluated at 10th and 90th percentiles of EA-PGI distribution. We also present main figure with categorical deciles of EA-PGI in Appendix Figure A.2, which is visually consistent with linear effect of EA-PGI. When computing cumulated lifetime income in the tables and text, we discount income by the a 3% interest rate, therefore computing its net present value upon graduation. In the computation, we sum over all 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 (10th and 90th percentiles).

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

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 $t = 1, ..., 25$.

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).[10] 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 51 736 individuals and 972 917 person-year observations (see Table A.1).

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

References

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.

Almond, Douglas and Janet Currie (2011). “Killing Me Softly: The Fetal Origins Hypothesis”. *Journal of Economic Perspectives* 25.3, 153–172.

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.

Black, Sandra E., Neil Duzett, Adriana Lleras-Muney, Nolan Pope, and Joseph Price (2024). “Intergenerational Transmission of Lifespan in the US”. Mimeo.

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.

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.

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.

Ghirardi, Gaia, Carlos J Gil-Hernández, Fabrizio Bernardi, Elsje Van Bergen, and Perline Demange (2024). "Interaction of family SES with children's genetic propensity for cognitive and noncognitive skills: No evidence of the Scarr-Rowe hypothesis for educational outcomes". *Research in Social Stratification and Mobility* 92, 100960.

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.

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.

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.

Pereira, Rita Dias, Pietro Biroli, Titus Galama, Cornelius A Rietveld, and Kevin Thom (2022). "Gene–Environment Interplay in the Social Sciences". arXiv 2203.02198.

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.

Rimfeld, Kaili, Eva Krapohl, Maciej Trzaskowski, Jonathan RI Coleman, Saskia Selzam, Philip S Dale, Tonu Esko, Andres Metspalu, and Robert Plomin (2018). "Genetic influence on social outcomes during and after the Soviet era in Estonia". *Nature human behaviour* 2.4, 269–275.

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.

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

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.

Von Wachter, Till (2020). "The Persistent Effects of Initial Labor Market Conditions for Young Adults and Their Sources". *Journal of Economic Perspectives* 34.4, 168–194.

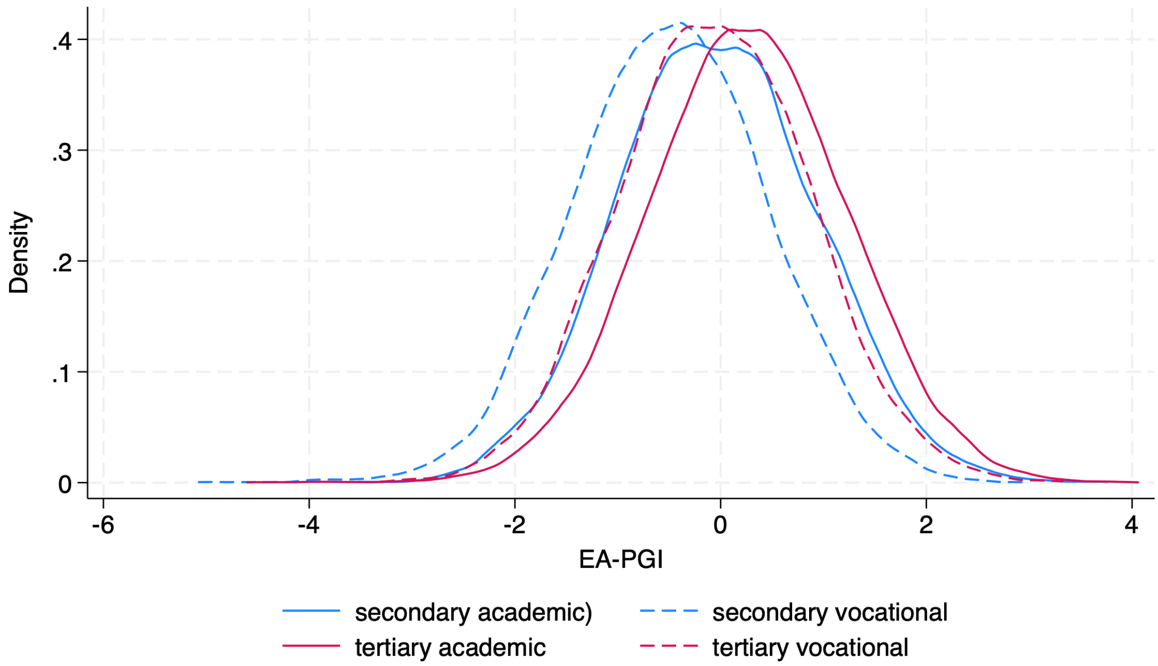
Wang, Biyao, Jessie R Baldwin, Tabea Schoeler, Rosa Cheesman, Wikus Barkhuizen, Frank Dudbridge, David Bann, Tim T Morris, and Jean-Baptiste Pingault (2021). "Robust genetic nurture effects on education: A systematic review and meta-analysis based on 38,654 families across 8 cohorts". *The American Journal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

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 Appendix

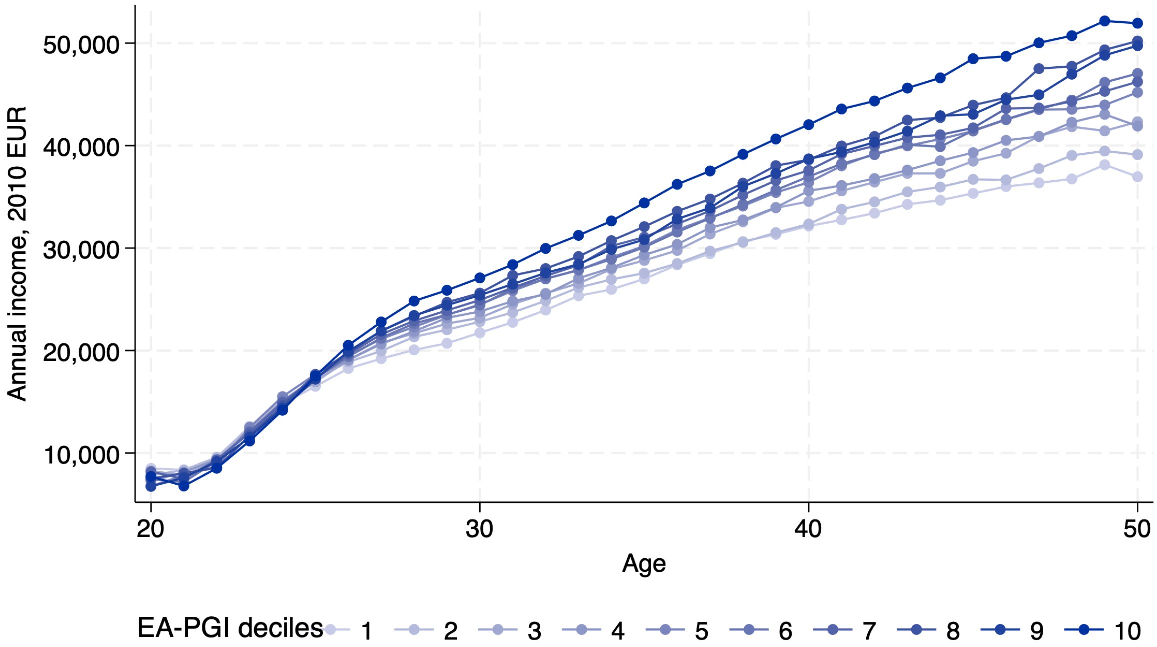
A.1 Additional figures

Figure A.1: Density of the EA-PGI by highest education level and track



Notes: density plot of EA-PGI in the working sample by highest education level and track.

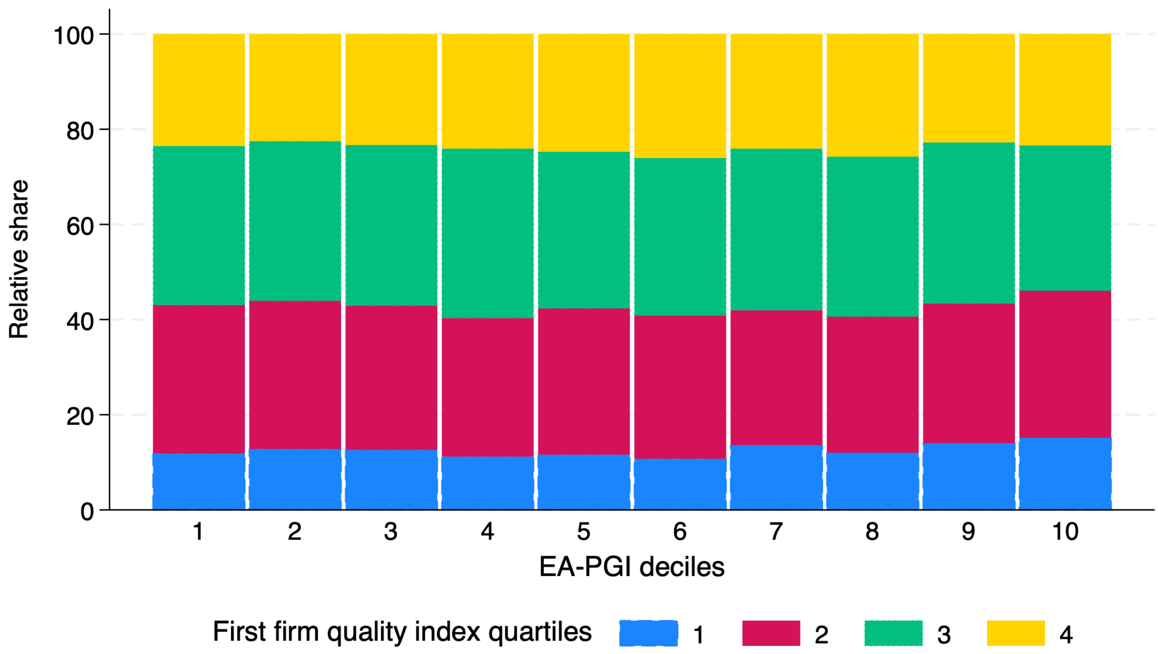
Figure A.2: Average earnings post graduation for tertiary educated by EA-PGI deciles



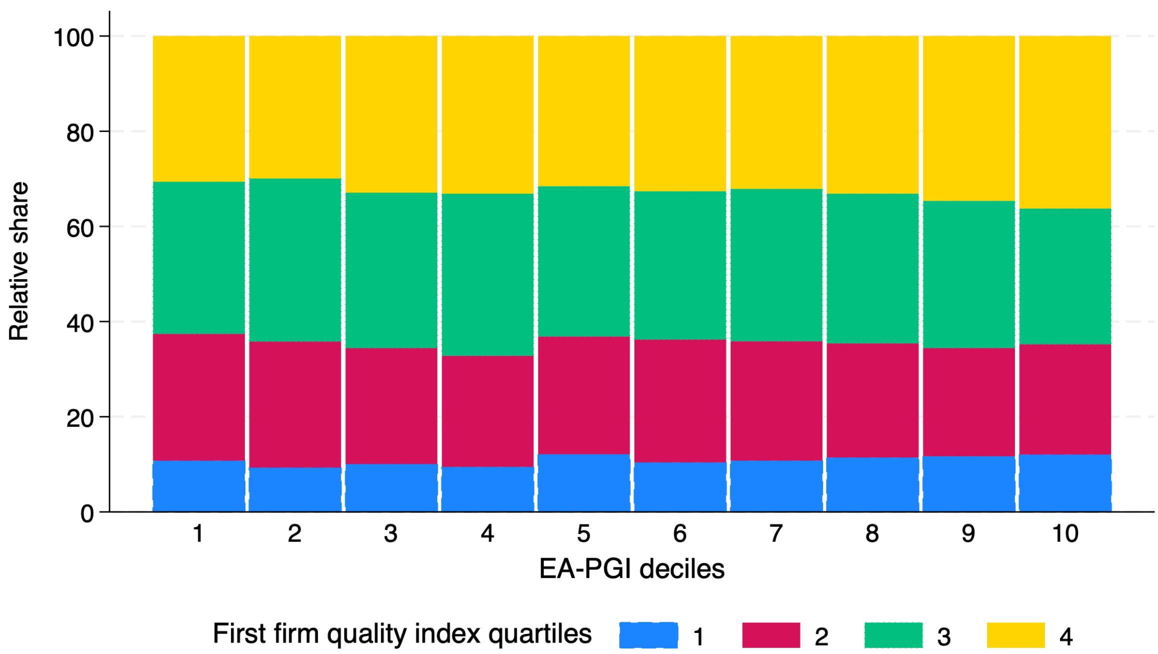
Notes: the figure plots predictive margins of average annual income (in 2010 EUR) by age. The different lines correspond to deciles of the EA-PGI. Average income is estimated following the regression of average annual income on EA-PGI deciles fully interacted with indicators measuring years since graduation and controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. The shaded areas correspond to 95% CI.

Figure A.3: Share of entry workers in each quartile of firm quality, by EA-PGI decile

(a) Secondary

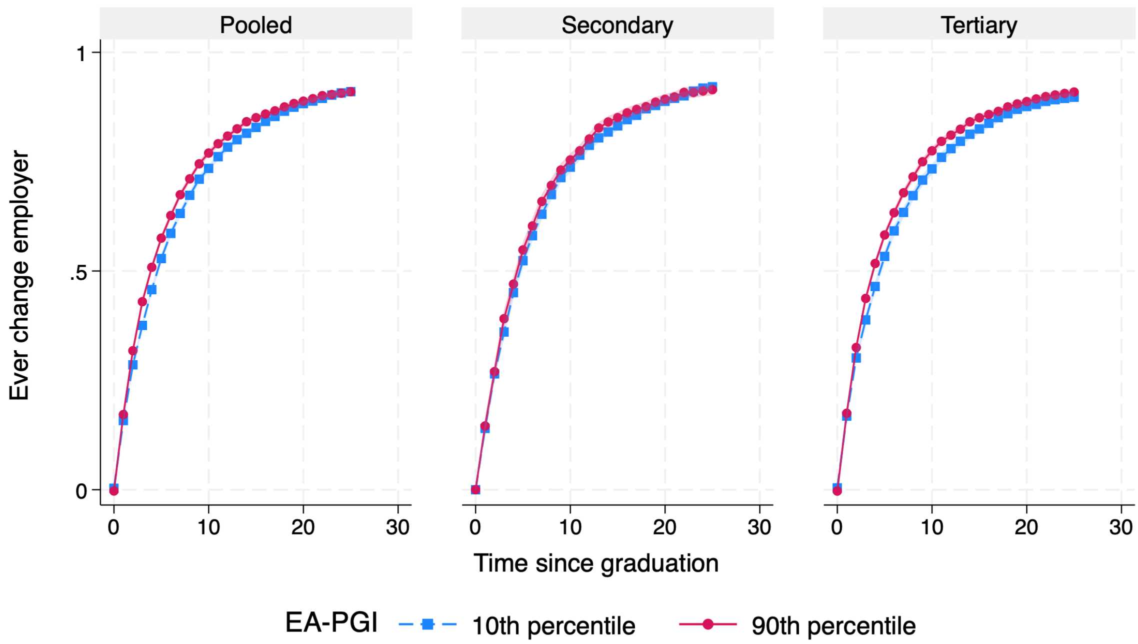


(b) Tertiary



Notes: the figure plots shares of workers by quartiles of the firm quality index of the first firm in individual's employment history. The shares are plotted across deciles of EA-PGI distribution. Panels A and B report the relative shares among secondary- and tertiary-educated workers, respectively.

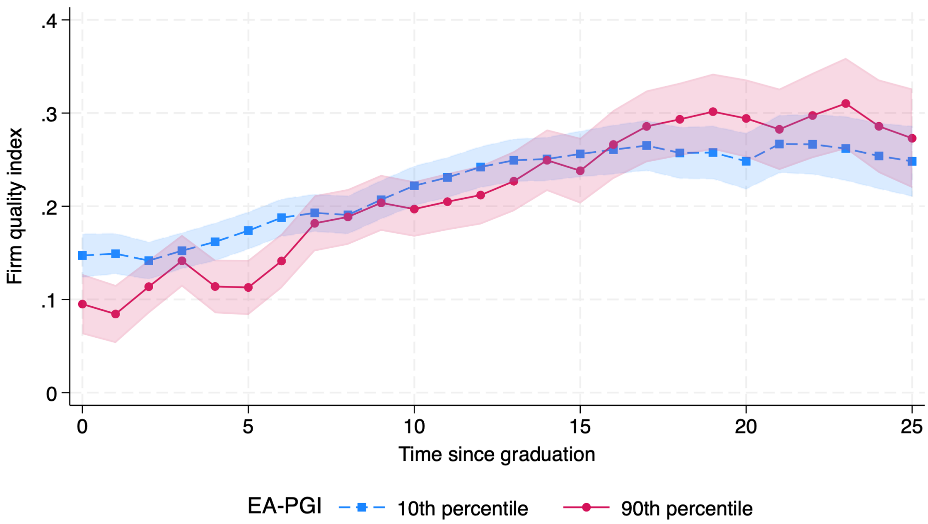
Figure A.4: Any change in employer since first job



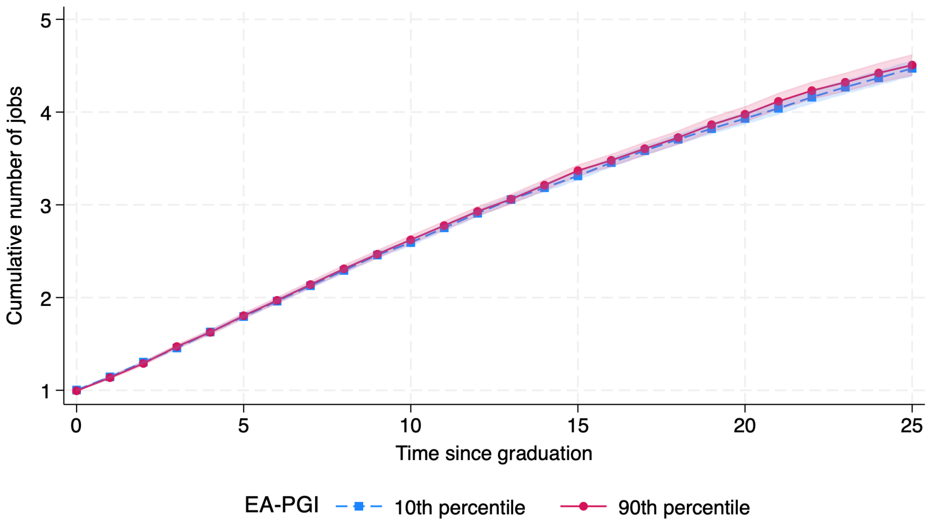
Notes: the figure plots average share of workers that switch employers at least once over time since graduation. The lines correspond to 10th and 90th EA-PGI percentiles. The first panel shows the average mobility patterns in the full sample, while other panels are restricted to secondary- and tertiary-educated workers, respectively. The average shares are computed from regressions of the indicator of having ever swtiched an employer 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.5: Employer quality and number of jobs since labor market entry among secondary-educated workers

(a) Firm quality since labor market entry



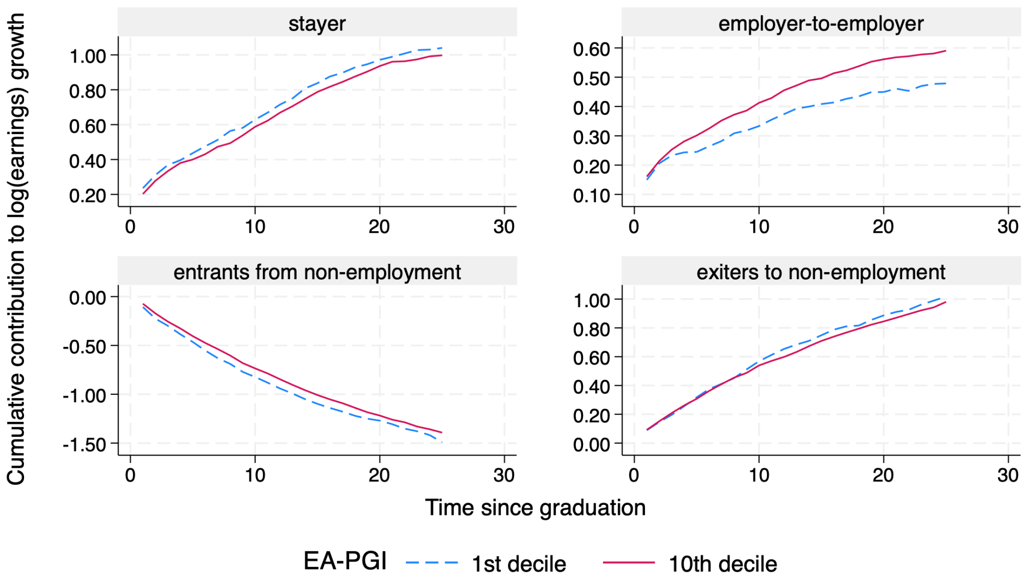
(b) Number of jobs since labor market entry



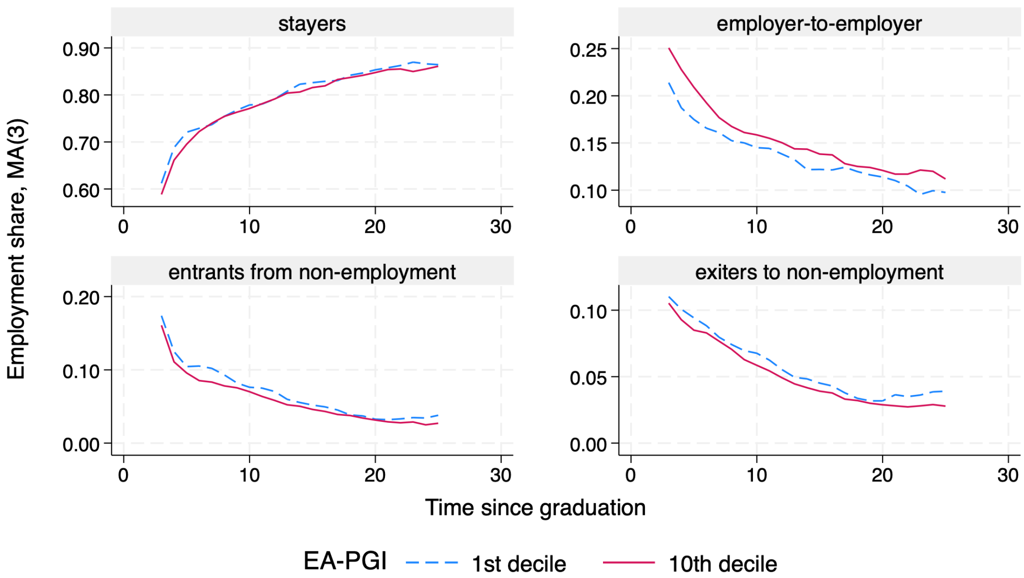
Notes: Panel A plots the average firm quality index over time since gradution at 10th and 90th percentiles of EA-PGI. Panel B plots the average number of employment spells over time since graduation at 10th and 90th percentiles of EA-PGI. Both are estimated from a regression of respective outcome on 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. The shaded areas correspond to 95% CIs.

Figure A.6: Decomposition of total cumulative growth in log(earnings) and group-specific employment shares

(a) Cumulative contribution to log(earnings) growth



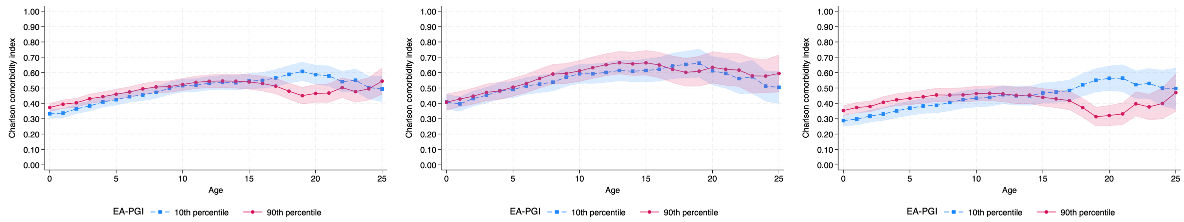
(b) Employment share



Notes: the figure reports the results of the decomposition of total earnings growth to within-firm growth, employer-to-employer switches, and mobility to/from non-employment. Panel A reports the cumulative contributions of each channel to to the total earnings growth over time since graduation at first and tenth EA-PGI decile. Panel B reports average employment shares of each type of workers over time and by EA-PGI deciles. The decomposition is applied to log annual earnings and follows the algorithm in Hahn, Hyatt, and Janicki, 2021. The sample for the decomposition is restricted to tertiary-educated workers.

Figure A.7: Average health indices by EA-PGI percentiles over time and by education after controlling for parental EA-PGI

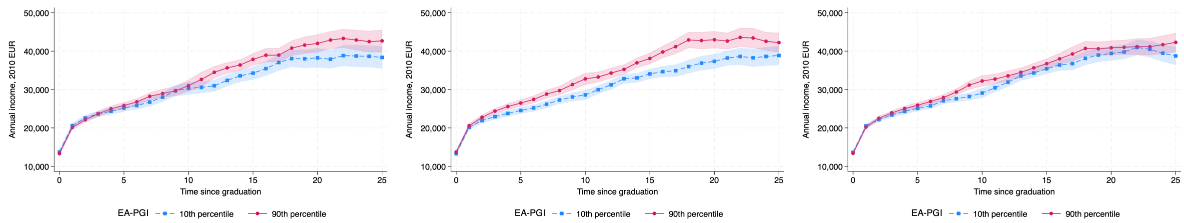
(a) Pooled (b) Secondary (c) Tertiary



Notes: the figure plots average health indices over time since graduation. The lines correspond to 10th and 90th percentiles of offspring EA-PGI. Panel A uses full sample of workers, while Panels B and C are restricted to secondary- and tertiary-educated workers, respectively. The estimates are obtained from a regression of Charlson Comorbidity Index on offspring, paternal and maternal 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 12 918 parent-offspring trios (directly genotyped and imputed). The shaded areas correspond to 95% CIs.

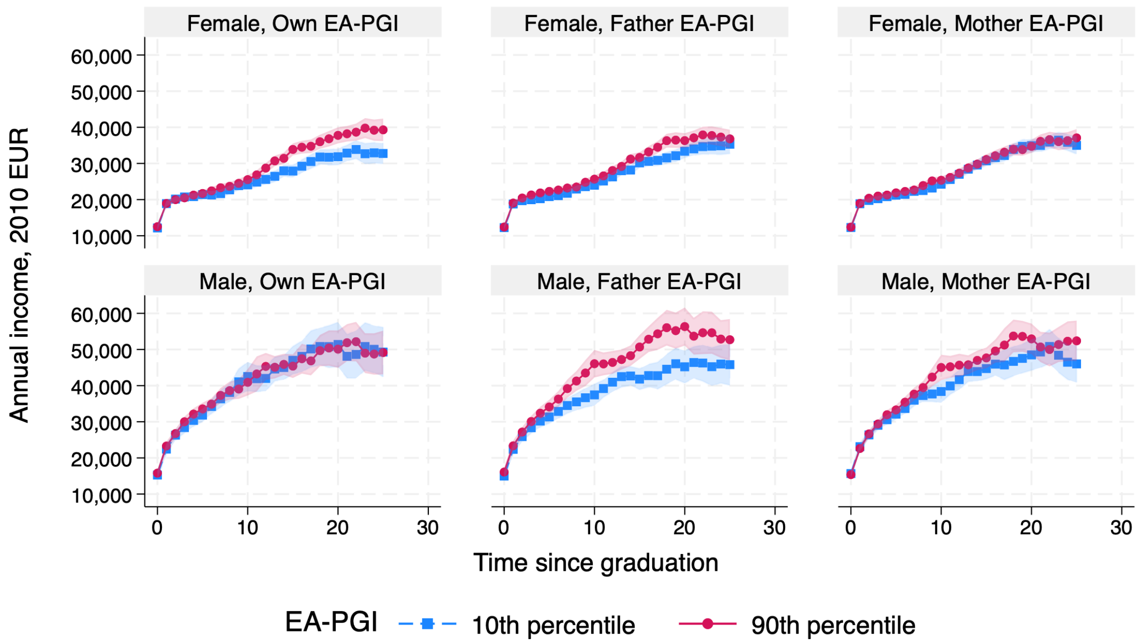
Figure A.8: Average annual income by own or parental EA-PGI percentiles over time

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



Notes: the figure plots average annual income over time since graduation. 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; in panel C - 10th and 90th percentiles of maternal EA-PGI. The estimates are computed from regression of annual income on own, paternal and maternal 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. The shaded areas correspond to 95% CIs.

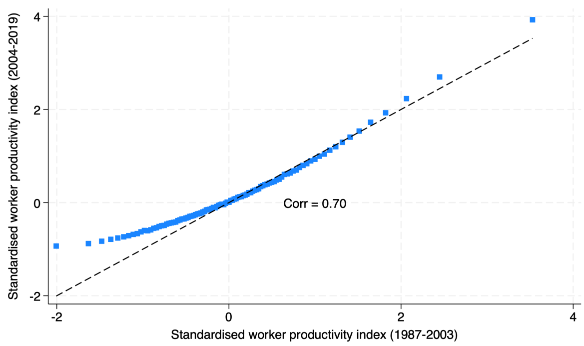
Figure A.9: Average annual income by own or parental EA-PGI percentiles and gender



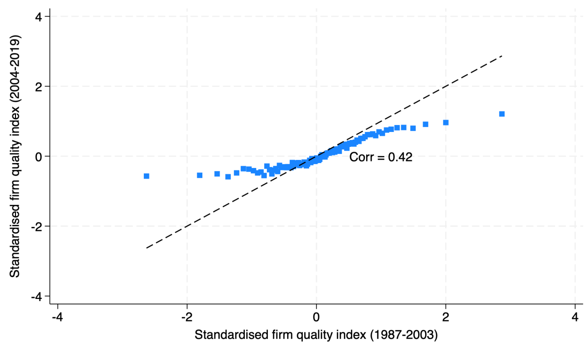
Notes: the figure plots average income evaluated at 10th and 90th percentiles of either own or parents' EA-PGI by index person's gender over time. The figure is based on the regression of annual income on own and parents' EA-PGI fully interacted with time and gender. The estimation sample is restricted to tertiary-educated workers. The estimation additionally controls for first ten genetic principal components, year of birth, calendar year and biobank indicators. The shaded areas correspond to 95% CI.

Figure A.10: Binscatter plot of AKM worker productivity and firm quality indices between two estimation periods (1987-2003 and 2004-2019)

(a) Worker productivity indices



(b) Firm quality indices



Notes: 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). Similarly, 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 workers observed in both periods; in Panel B - firms observed in both periods.

A.2 Additional tables

Table A.1: Working sample in trajectory analysis

	Person-year observations			Unique individuals		
	All	THL	BDB	All	THL	BDB
Start	5,374,521	3,963,254	1,411,267	176,523	132,171	44,352
Keep graduates only	3,248,655	1,957,308	1,291,347	100,016	59,416	40,600
Keep graduates with non-missing graduation year	3,248,655	1,957,308	1,291,347	100,016	59,416	40,600
Graduated between 1970 and 2020	3,173,023	1,882,118	1,290,905	97,390	56,803	40,587
Observed between 0 and 25 years since graduation	1,707,370	1,065,113	642,257	97,369	56,786	40,583
Followed from 0 years since graduation	1,011,416	576,894	434,522	58,904	29,205	29,699
Followed at least up to age 30 (if secondary)	972,897	567,236	405,661	51,735	27,135	24,600

Table A.2: Average population and sample characteristics of graduates

(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	1,599,341	51,735
Cohort: 1960-69	0.17	0.22	0.16	0.000	1.000	1,599,341	51,735
Cohort: 1970-79	0.34	0.36	0.34	0.000	1.000	1,599,341	51,735
Cohort: 1980-89	0.36	0.29	0.37	0.000	1.000	1,599,341	51,735
Cohort: 1990-99	0.13	0.12	0.13	0.000	1.000	1,599,341	51,735
Graduation age: 16-20	0.36	0.31	0.36	0.000	1.000	1,599,341	51,735
Graduation age: 21-25	0.39	0.43	0.39	0.000	1.000	1,599,341	51,735
Graduation age: 26-30	0.25	0.25	0.24	0.001	1.000	1,599,341	51,735
Education: secondary	0.44	0.37	0.44	0.000	1.000	1,599,341	51,735
Education: tertiary	0.56	0.63	0.56	0.000	1.000	1,599,341	51,735
Male	0.48	0.39	0.48	0.000	1.000	1,599,341	51,735
Married	0.10	0.13	0.11	0.000	0.000	1,599,341	51,735
Rural	0.24	0.24	0.25	0.966	1.000	1,599,341	51,735
Income at t=0	9,301	9,523	9,327	0.000	1.000	1,599,341	51,735

Notes: Average characteristics measured in year of graduation in the population of graduates (1) and sample of genotyped graduates used in the analysis (2). The population of graduates is selected with similar criteria described in Table A.1. Column (3) shows average characteristics after having rebalanced 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. Column (4) reports p-values for the equality of means between population and genotyped sample, while column (5) does so for the population vs. reweighted sample comparison. All p-values are adjusted for multiple hypotheses testing via Holm correction. Columns (6) and (7) report population and genotyped sample counts, respectively.

(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	0.830	1.000	1,599,341	12,918
Cohort: 1960-69	0.17	0.15	0.17	0.002	1.000	1,599,341	12,918
Cohort: 1970-79	0.34	0.35	0.34	0.048	1.000	1,599,341	12,918
Cohort: 1980-89	0.36	0.38	0.36	0.000	1.000	1,599,341	12,918
Cohort: 1990-99	0.13	0.11	0.12	0.000	1.000	1,599,341	12,918
Graduation age: 16-20	0.36	0.34	0.36	0.000	1.000	1,599,341	12,918
Graduation age: 21-25	0.39	0.42	0.40	0.000	1.000	1,599,341	12,918
Graduation age: 26-30	0.25	0.24	0.24	0.132	1.000	1,599,341	12,918
Education: secondary	0.44	0.39	0.44	0.000	1.000	1,599,341	12,918
Education: tertiary	0.56	0.61	0.56	0.000	1.000	1,599,341	12,918
Male	0.48	0.40	0.48	0.000	1.000	1,599,341	12,918
Married	0.10	0.12	0.12	0.000	1.000	1,599,341	12,918
Rural	0.24	0.27	0.24	0.000	1.000	1,599,341	12,918
Income at t=0	9,301	9,788	9,560	0.000	1.000	1,599,341	12,918

Notes: Average characteristics measured in year of graduation in the population of graduates (1) and sample of genotyped family trio graduates used in the analysis (2). The population of graduates is selected with similar criteria described in Table A.1. Column (3) shows average characteristics after having rebalanced 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. Column (4) reports p-values for the equality of means between population and genotyped family trio sample, while column (5) does so for the population vs. reweighted sample comparison. All p-values are adjusted for multiple hypotheses testing via Holm correction. Columns (6) and (7) report population and genotyped family trio sample counts, respectively.

Table A.3: Summary statistics in FOLK-based AKM sample

	Mean	SD	N
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
psi	0.365	0.771	15,699,075
theta	0.122	0.990	16,162,338

Notes: the table reports summary statistics in FOLK-based matched employee-employer panel used in AKM estimation.

Table A.4: AKM summary statistics and variance decomposition

	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

Notes: the table reports summary statistics and variance decomposition following AKM estimations in FOLK-based matched employee-employer panel. The dependent variable 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 yearly median, working in firms with at least 10 workers and for at least 4 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.

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

	Sample means			Diff.				
	1st decile	2nd-9th deciles	10th decile	2nd-9th deciles		10th decile		
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

Notes: the table reports descriptive statistics among tertiary-educated individuals observed in trajectory panel by deciles of EA-PGI. The first three columns report estimated sample means and standard error of mean in parentheses in each EA-PGI decile group. The next two columns report estimated differences of sample means in 2nd–9th and 10th deciles relative to 1st decile, as well as standard error of difference estimate in parentheses. The last column reports total sample count for each variable considered. All estimates control for first 10 genetic PCs. All estimates are unweighted.

Table A.6: Worker productivity and EA-PGI

	Dependent variable: Std AKM worker FE			
	(1)	(2)	(3)	(4)
Compulsory × Std EA-PGI	0.044***	0.044***	0.044***	0.044***
	(0.007)	(0.006)	(0.006)	(0.006)
Secondary × Std EA-PGI	0.022***	0.017***	0.015**	0.012*
	(0.005)	(0.005)	(0.005)	(0.005)
Tertiary × Std EA-PGI	0.166***	0.111***	0.105***	0.052***
	(0.005)	(0.005)	(0.005)	(0.005)
Level	Yes	Yes	Yes	Yes
Field	No	Yes	Yes	Yes
Level × Field	No	No	Yes	No
Institution ID	No	No	No	Yes
Obs.	84,854	73,451	73,447	73,307
Avg. obs. per cell	28284.667	734.510	292.618	46.633
Adj R2	0.056	0.095	0.103	0.154
RMSE	0.852	0.778	0.774	0.752

Standard errors in parentheses

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

Notes: the table reports estimation results of worker FEs on EA-PGI of workers by education level. All regressions control for first 10 genetic PCs, gender, year of birth and calendar year indicators. Furthermore, the columns gradually add controls for education level, field and institution ID, as well as their interactions. All estimations are unweighted. Standard errors are reported in parentheses.

Table A.7: Years of education and EA-PGI with and without parental EA-PGI

	Dependent var: predicted years of education	
	All family trios	Directly genotyped trios
Baseline without parental EA-PGI		
Own EA-PGI	0.283***	0.274***
	(0.011)	(0.019)
Constant	14.496***	13.836***
	(0.505)	(1.308)
Obs.	12 871	4 586
Controlling for parental EA-PGI		
Own EA-PGI	0.212***	0.224***
	(0.017)	(0.026)
Mother EA-PGI	0.056***	0.048**
	(0.014)	(0.019)
Father EA-PGI	0.057***	0.032*
	(0.013)	(0.019)
Constant	14.467***	13.835***
	(0.501)	(1.294)
Obs.	12 871	4 586

$^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$

Notes: the table reports estimation results of predicted years of education (given highest qualification) on EA-PGI. The top panel reports baseline estimations without controlling for parental EA-PGI. The bottom panel reports results from estimations controlling for parental EA-PGI. All estimations additionally control for first ten genetic principal components, gender, year of birth, calendar and biobank indicators. The results in second column were obtained in the sample of all trios (including those whose parents' genotypes were imputed from incomplete trios or siblings). In the third column, the estimation sample is restricted to only the trios that were directly genotyped. Standard errors reported in parentheses.

Table A.8: Cumulated lifetime income by own and parents' EA-PGI percentiles and gender

	Dependent variable: cumulated income					
	Own EA-PGI		Father EA-PGI		Mother EA-PGI	
	Male	Female	Male	Female	Male	Female
10th percentile	441 954	285 287	417 298	286 268	434 859	290 561
	(12 946)	(5 277)	(10 162)	(4 107)	(9 838)	(4 208)
90th percentile	444 162	304 811	469 661	304 026	451 235	299 437
	(11 618)	(5 116)	(11 931)	(4 491)	(9 939)	(4 353)
Obs.	2 663	5 145	2 663	5 145	2 663	5 145

Notes: the table reports adjusted lifetime income (up to 25 years since graduation) at 10th and 90th percentiles of EA-PGI distribution. Average income adjusted by regressing cumulated income on parental EA-PGI fully interacted with index person's gender, first ten genetic principal components, year of birth, calendar, and biobank indicators. The estimation sample is restricted to tertiary educated index people. Income discounted to obtain its net present value upon graduation (see Section 4.2 for additional information). Standard errors reported in parentheses.

Table A.9: Cumulated lifetime income by own EA-PGI percentiles: weighted and unweighted analysis

	Dependent variable: Cumulated income					
	Pooled		Secondary		Tertiary	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
Panel A: Genotyped sample (baseline specification)						
10th percentile	309 438	291 728	262 462	257 996	345 947	331 362
	(1 316)	(1 286)	(1 440)	(1 501)	(1 959)	(1 920)
50th percentile	329 829	308 756	255 444	249 549	368 656	350 947
	(856)	(832)	(1 114)	(1 157)	(1 137)	(1 105)
90th percentile	350 395	325 930	248 366	241 029	391 560	370 700
	(1 590)	(1 525)	(2 119)	(2 185)	(2 004)	(1 938)
Obs.	51 056	51 056	18 692	18 692	32 364	32 364
Panel B: Family trio sample (controlling for parents)						
10th percentile	303 711	304 546	259 965	263 787	339 170	345 745
	(3 910)	(4 417)	(4 625)	(5 261)	(5 695)	(6 559)
50th percentile	313 178	313 886	255 342	258 422	345 629	351 212
	(1 697)	(1 981)	(2 153)	(2 388)	(2 321)	(2 755)
90th percentile	322 768	323 347	250 658	252 986	352 173	356 750
	(3 936)	(4 502)	(5 297)	(5 717)	(5 201)	(6 114)
Obs.	12 871	12 871	5 063	5 063	7 808	7 808

Notes: The table reports adjusted lifetime income (up to 25 years since graduation) by own EA-PGI percentiles. Panel A reports the baseline results obtained from regression of cumulated income on EA-PGI controlling for first ten genetic principal components, gender, year of birth, calendar year, and biobank indicators. Panel B reports the results in the family trio subsample obtained from similar regression, but additionally controlling for parents' EA-PGI as well. Income discounted to obtain its net present value upon graduation (see Section 4.2 for additional information). The weighted analysis use inverse probability weights that we compute to match average birth cohort, education, graduation year, gender and rural residence indicators in the population of Finnish graduates. Standard errors reported in parentheses.