Genes and peers: a review of empirical applications

Giorgia Menta

Luxembourg Institute of Socio-Economic Research (LISER)

18^{th} Winter School on Inequality and Social Welfare Theory

January 7^{th}, 2025

Peer effects in economics

- Whom we interact with affects how we behave and perform
- Peer effects arise whenever economic agents interact within a group or a network
- **Problem 1**: individuals sort endogenously into peer groups (Fletcher, 2012)
  - Friends
  - Co-workers
  - School mates
- **Problem 2**: even if sorting was exogenous, behavioural responses and homophily are likely not!
  - ITT interpretation

- Some assumptions are needed to bring in identification

Peer effects in economics

- Whom we interact with affects how we behave and perform
- Peer effects arise whenever economic agents interact within a group or a network
- **Problem 1**: individuals sort endogenously into peer groups (Fletcher, 2012)
  - Friends
  - Co-workers
  - School mates
- **Problem 2**: even if sorting was exogenous, behavioural responses and homophily are likely not!
  - ITT interpretation
- Some assumptions are needed to bring in identification

Peer effects in economics

- Whom we interact with affects how we behave and perform
- Peer effects arise whenever economic agents interact within a group or a network
- **Problem 1**: individuals sort endogenously into peer groups (Fletcher, 2012)
  - Friends
  - Co-workers
  - School mates
- **Problem 2**: even if sorting was exogenous, behavioural responses and homophily are likely not!
  - ITT interpretation

- Some assumptions are needed to bring in identification

Peer effects in economics

- Whom we interact with affects how we behave and perform
- Peer effects arise whenever economic agents interact within a group or a network
- **Problem 1**: individuals sort endogenously into peer groups (Fletcher, 2012)
  - Friends
  - Co-workers
  - School mates
- **Problem 2**: even if sorting was exogenous, behavioural responses and homophily are likely not!
  - ITT interpretation

- Some assumptions are needed to bring in identification

Addressing endogenous sorting

- **Random assignment**:
  - Carrell, Sacerdote & West (2013); Booij, Leuven & Oosterbeek (2017); Oosterbeek & van Ewijk (2014)
- **RDD**:
  - Test-scores (Pop-Eleches & Urquiola, 2013; Varadottir, 2013), university admission scores (Ribas, Sampaio, & Trevisan, 2024), wages (Dube, Giuliano, & Leonard, 2019), school-boundaries (Billings, Deming & Rockoff, 2014), date of birth (Dustmann, Puhani and Schönberg, 2017)
- **IV**:
  - Peers-of-peers' characteristics (e.g. Bramoullé, Djebbari and Fortin, 2009; de Giorgi, Pellizzari and Redaelli, 2010; Mendolia, Paloyo and Walker, 2018)

Addressing endogenous sorting (continued)

- **Fixed effects**:
  - Most popular tool
  - Identification: random assignment on lowest levels of disaggregation (e.g. school-grade peers within a given school and school-year)
  - Large literature!
- Evidence from this literature holds important implications for the design of tracking, grouping, and admission systems

Segregation vs mixing

- E.g. **Education**: should we segregate students based on their academic performance?
  - Depends on whether we see a child's ability as a *substitute* or a *complement* for her peers' ability
    - *Complements*: segregation ('bad apple' theory: Lazear, 2001)
    - *Substitutes*: mixing ('good apple' theory)
  - Role of *market failures* (e.g. credit constraints): we cannot deduce complementarity from observed sorting
    - e.g. rich parents sending rich kids to the best/more expensive schools
  - 'Technological' vs 'sociological' effects
- Mixed empirical evidence, although most applications suggest that school sorting/tracking is generally bad

Segregation vs mixing

- E.g. **Education**: should we segregate students based on their academic performance?
  - Depends on whether we see a child's ability as a *substitute* or a *complement* for her peers' ability
    - *Complements*: segregation ('bad apple' theory: Lazear, 2001)
    - *Substitutes*: mixing ('good apple' theory)
  - Role of *market failures* (e.g. credit constraints): we cannot deduce complementarity from observed sorting
    - e.g. rich parents sending rich kids to the best/more expensive schools
  - 'Technological' vs 'sociological' effects
- Mixed empirical evidence, although most applications suggest that school sorting/tracking is generally bad

Segregation vs mixing

- E.g. **Education**: should we segregate students based on their academic performance?
  - Depends on whether we see a child's ability as a *substitute* or a *complement* for her peers' ability
    - *Complements*: segregation ('bad apple' theory: Lazear, 2001)
    - *Substitutes*: mixing ('good apple' theory)
  - Role of *market failures* (e.g. credit constraints): we cannot deduce complementarity from observed sorting
    - e.g. rich parents sending rich kids to the best/more expensive schools
  - 'Technological' vs 'sociological' effects
- Mixed empirical evidence, although most applications suggest that school sorting/tracking is generally bad

Segregation vs mixing

- E.g. **Education**: should we segregate students based on their academic performance?
  - Depends on whether we see a child's ability as a *substitute* or a *complement* for her peers' ability
    - *Complements*: segregation ('bad apple' theory: Lazear, 2001)
    - *Substitutes*: mixing ('good apple' theory)
  - Role of *market failures* (e.g. credit constraints): we cannot deduce complementarity from observed sorting
    - e.g. rich parents sending rich kids to the best/more expensive schools
  - 'Technological' vs 'sociological' effects
- Mixed empirical evidence, although most applications suggest that school sorting/tracking is generally bad

Why genetics

Why bringing genetics into the study of peer effects?

- Pros:
  - Conditional exogeneity if able to control for parental genotypes
  - 'Invisible' sorting based on genotypes arguably less strong than 'visible' sorting based on SES
- Cons:
  - Currently challenging to isolate direct genetic effects
  - More distal (and likely smaller) effects than literature on peers' phenotypes

What will we talk about today?

- Review the literature on genetic peer effects
  - Siblings vs school-grade mates
- Present some preliminary results from a new paper

- **Notes**: All papers here rely on population-based GWAS. From Biroli et al. (2022):
  - If within-family G: *downward-biased* estimates of G (overcontrolling for direct genetic effects)
    - Can be corrected, e.g. Becker et al. (2021)
  - If no family data: *upward-biased* estimates of G (may reflect parental G or environmental confounding)

What will we talk about today?

- Review the literature on genetic peer effects
  - Siblings vs school-grade mates
- Present some preliminary results from a new paper

- **Notes**: All papers here rely on population-based GWAS. From Biroli et al. (2022):
  - If within-family G: *downward-biased* estimates of G (overcontrolling for direct genetic effects)
    - Can be corrected, e.g. Becker et al. (2021)
  - If no family data: *upward-biased* estimates of G (may reflect parental G or environmental confounding)

Siblings as peers

- The literature on peer effects of siblings is somewhat scarce:
  - Mostly focusing on spillover effects of a sibling's peer composition
  - Fewer studies looking at the causal effect of siblings characteristics
    - e.g. gender (Peter et al., 2018), fertility (Lyngstad & Prskawetz, 2010), parental leave take-up (Dahl, Løken & Mogstad, 2014)
- Genes can provide an ideal source of (conditional) quasi-experimental variation
  - Rauscher, Conley, & Siegal (2015); Cawley et al. (2023)

Siblings genetic endowments and parental investments

Sanz-de-Galdeano, A., & Terskaya, A. (2023). Sibling Differences in Genetic Propensity for Education: How do Parents React? *Review of Economics and Statistics*, forthcoming.

- **Dataset**: Add Health siblings pair (≈3,000 pairs)
- PGI for educational attainment
  - From PGI repository of Becker et al. (2021), which use weights from the EA3 GWAS (Lee et al., 2018)
- **Outcome**: parental investments (average of binary indicators)
- **Peer group**: siblings
- Within-family analysis: Mendelian imputation of parental genotypes (Young et al., 2022)

Theoretical framework

- **Parental utility function**: $U = U(c, V_1, ..., V_n)$
  - V_i is child i's human capital
  - Assume separability in consumption and a CES functional form:

$$U = [V_1^{\rho} + V_2^{\rho}]^{\frac{1}{\rho}}$$

- **Human capital production function**: $V(e_i, PI_i) = e_i^{\alpha_e} PI_i^{\alpha_p}$
  - e_i is child i's genetic endowment
  - e_i and PI_i are complements
- S.t. budget constraint, then it can be shown that:

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} < 0 \iff \rho < 0 \text{ (equality} \succ \text{efficiency)}$$

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} > 0 \iff 0 < \rho < 1 \text{ (efficiency} \succ \text{equality)}$$

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} = 0 \iff \rho = 1 \text{ (efficiency-equality trade-off)}$$

Theoretical framework

- **Parental utility function**: $U = U(c, V_1, ..., V_n)$
  - V_i is child i's human capital
  - Assume separability in consumption and a CES functional form:

$$U = [V_1^\rho + V_2^\rho]^{\frac{1}{\rho}}$$

- **Human capital production function**: $V(e_i, PI_i) = e_i^{\alpha_e} PI_i^{\alpha_p}$
  - e_i is child i's genetic endowment
  - e_i and PI_i are complements
- S.t. budget constraint, then it can be shown that:

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} < 0 \iff \rho < 0 \text{ (equality} \succ \text{efficiency)}$$

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} > 0 \iff 0 < \rho < 1 \text{ (efficiency} \succ \text{equality)}$$

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} = 0 \iff \rho = 1 \text{ (efficiency-equality trade-off)}$$

Theoretical framework

- **Parental utility function**: $U = U(c, V_1, ..., V_n)$
  - V_i is child i's human capital
  - Assume separability in consumption and a CES functional form:

$$U = [V_1^{\rho} + V_2^{\rho}]^{\frac{1}{\rho}}$$

- **Human capital production function**: $V(e_i, PI_i) = e_i^{\alpha_e} PI_i^{\alpha_p}$
  - e_i is child i's genetic endowment
  - e_i and PI_i are complements
- S.t. budget constraint, then it can be shown that:

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} < 0 \iff \rho < 0 \text{ (equality} \succ \text{efficiency)}$$

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} > 0 \iff 0 < \rho < 1 \text{ (efficiency} \succ \text{equality)}$$

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} = 0 \iff \rho = 1 \text{ (efficiency-equality trade-off)}$$

Theoretical framework

- **Parental utility function**: $U = U(c, V_1, ..., V_n)$
  - V_i is child i's human capital
  - Assume separability in consumption and a CES functional form:

$$U = [V_1^\rho + V_2^\rho]^{\frac{1}{\rho}}$$

- **Human capital production function**: $V(e_i, PI_i) = e_i^{\alpha_e} PI_i^{\alpha_p}$
  - e_i is child i's genetic endowment
  - e_i and PI_i are complements
- S.t. budget constraint, then it can be shown that:

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} < 0 \iff \rho < 0 \text{ (equality} \succ \text{efficiency)}$$

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} > 0 \iff 0 < \rho < 1 \text{ (efficiency} \succ \text{equality)}$$

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} = 0 \iff \rho = 1 \text{ (efficiency-equality trade-off)}$$

Theoretical framework

- **Parental utility function**: $U = U(c, V_1, ..., V_n)$
  - V_i is child i's human capital
  - Assume separability in consumption and a CES functional form:

$$U = [V_1^{\rho} + V_2^{\rho}]^{\frac{1}{\rho}}$$

- **Human capital production function**: $V(e_i, PI_i) = e_i^{\alpha_e} PI_i^{\alpha_p}$
  - e_i is child i's genetic endowment
  - e_i and PI_i are complements
- S.t. budget constraint, then it can be shown that:

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} < 0 \iff \rho < 0 \text{ (equality} \succ \text{efficiency)}$$

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} > 0 \iff 0 < \rho < 1 \text{ (efficiency} \succ \text{equality)}$$

$$\frac{\partial log(PI_1|e_1)}{\partial(\frac{e_1}{e_2})} = 0 \iff \rho = 1 \text{ (efficiency-equality trade-off)}$$

Main results

Table 3: The Effect of Educational Polygenic Index and Sibling Differences in Educational Polygenic Indexes on Parental Investments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pooled Sample			Non-Twins			Twins		
EA PGI-Sibling's EA PGI	-0.232	-0.233	-0.226	-0.265	-0.270	-0.256	-0.050	-0.043	-0.052
SE	(0.084)	(0.081)	(0.082)	(0.101)	(0.100)	(0.102)	(0.134)	(0.129)	(0.138)
$p-value$	[0.006]	[0.004]	[0.006]	[0.009]	[0.007]	[0.012]	[0.708]	[0.737]	[0.710]
EA PGI	0.205	0.212	0.201	0.168	0.174	0.148	0.245	0.266	0.273
SE	(0.080)	(0.077)	(0.086)	(0.093)	(0.095)	(0.117)	(0.135)	(0.123)	(0.130)
$p-value$	[0.011]	[0.006]	[0.019]	[0.074]	[0.067]	[0.206]	[0.071]	[0.032]	[0.037]
Parental EA PGI	-0.001	-0.001	-0.002	0.011	0.023	0.008	-0.011	-0.014	-0.023
SE	(0.073)	(0.074)	(0.089)	(0.089)	(0.093)	(0.123)	(0.110)	(0.107)	(0.116)
$p-value$	[0.984]	[0.994]	[0.978]	[0.903]	[0.800]	[0.945]	[0.921]	[0.893]	[0.846]
N	604	604	604	412	412	412	192	192	192
Baseline controls	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
Additional controls	*No*	*No*	*Yes*	*No*	*No*	*Yes*	*No*	*No*	*Yes*

Note: EA PGI is the educational attainment polygenic index. The table reports the estimated effects of sibling differences in EA PGI, own EA PGI, and parental EA PGI on parental investments (as measured by an index standardized to have mean 0 and standard deviation 1) in the pooled sample 1-3), in the sample of non-twins 4-6, and in the sample of twins 7-9. EA PGI is always standardized to have mean 0 and standard deviation 1. The regressions in columns 2, 5, and 8 include age, age-squared, sibling differences in age (only included in the non-twins sample), a female dummy, and a female sibling dummy. The regressions in columns 3, 6, and 9 include in addition a rural area dummy, an indicator that both parents cohabit, and the SES index. Coefficient estimates and standard errors are measurement-error-corrected as described in Appendix F. Standard errors clustered at the family level are in parentheses.

Main results

- **Parents are inequality averse**: as the relative endowment of child 1 increases relative to child 2, they invest less in child 1 (keeping e_1 constant)
- Parental investments during adolescence can **attenuate** inequalities from the genetic lottery
- Effects are less strong for twins: difficulty to separate parental investments for children close in age (Bharadwaj et al., 2018; Terskaya, 2023)

School-mates as peers

- **Selection**: friends and schoolmates are more genetically similar to each other than randomly drawn individuals in the population (Domingue et al., 2018)
- **Social genetic effects/ genetic peer effects**: a peer's genotype affects my phenotype
- Domingue et al. (2018): Schoolmates and friends education PGI predicts own educational attainment (no social genetic effects for height or BMI)

Genetic peer effects on obesity

Brunello, G., Sanz-de-Galdeano, A., & Terskaya, A. (2020). Not only in my genes: The effects of peers' genotype on obesity. *Journal of Health Economics*, 72, 102349.

- **Dataset**: Add Health
- **PGI for BMI** (0.2 corr with BMI)
  - From PGI repository of Becker et al. (2021), which use weights from the Locke et al. (2015) BMI GWAS
- **Outcome**: Obesity
- **Peer group**: Same grade in school (grade-mates)

Main results: short-run (age 16)

Panel B: females				
Grademates BMI PGS (normalized)	0.028***	0.028***	0.028***	0.028***
	(0.009)	(0.009)	(0.009)	(0.010)
BMI PGS (normalized)	0.071***	0.062***	0.061***	0.060***
	(0.010)	(0.010)	(0.010)	(0.018)
Observations	2343	2343	2343	2343
R^2	0.149	0.176	0.177	0.180
Panel C: males				
Grademates BMI PGS (normalized)	0.004	0.001	0.000	−0.004
	(0.010)	(0.010)	(0.010)	(0.010)
BMI PGS (normalized)	0.059***	0.048***	0.049***	0.045***
	(0.010)	(0.009)	(0.009)	(0.017)
Observations	2107	2107	2107	2107
R^2	0.110	0.171	0.176	0.184
Individual controls	Yes	Yes	Yes	Yes
Family and parental controls		Yes	Yes	Yes
Principal components of SNP matrix			Yes	Yes
School-grade level controls				Yes

Note: The dependent variable is the obesity indicator measured at Wave II of Add Health. All regressions include school and grade fixed effects, and individual controls such as age, age squared, gender and race indicators. Family and parental controls (added in Columns 2–4) include family income, parental education, socio-economic status, parental obesity, the number of siblings and an indicator of whether both parents live in the household. We add in Column (3) the first 10 principal components of the full genetic data, and we further add in Column (4) average age, the share of females and blacks, grade size, average parental education—all computed at school-grade level—and average schoolmates BMI PGS computed at the school level. Standard errors (in parentheses) are clustered at the school level. Estimates are weighted using Wave I weights. ***$p<0.01$, **$p<0.05$, *$p<0.1$.

- +1 SD in peers BMI PGI increases P(obesity) by 2.8 percentage points for girls

Other results

- **Quantile regression**: effects on females mostly found on the top quartile of the BMI distribution
- **No long-run effects**: social-genetic effect fades over time (no longer different from zero 13 years after)
- **Mechanisms**: changed perception of body size (more likely to underestimate weight) and worse eating habits

- **Implications**: if there is endogenous sorting, randomly allocating pupils to classes might improve girls' outcomes (but only in the short term)

Other results

- **Quantile regression**: effects on females mostly found on the top quartile of the BMI distribution
- **No long-run effects**: social-genetic effect fades over time (no longer different from zero 13 years after)
- **Mechanisms**: changed perception of body size (more likely to underestimate weight) and worse eating habits

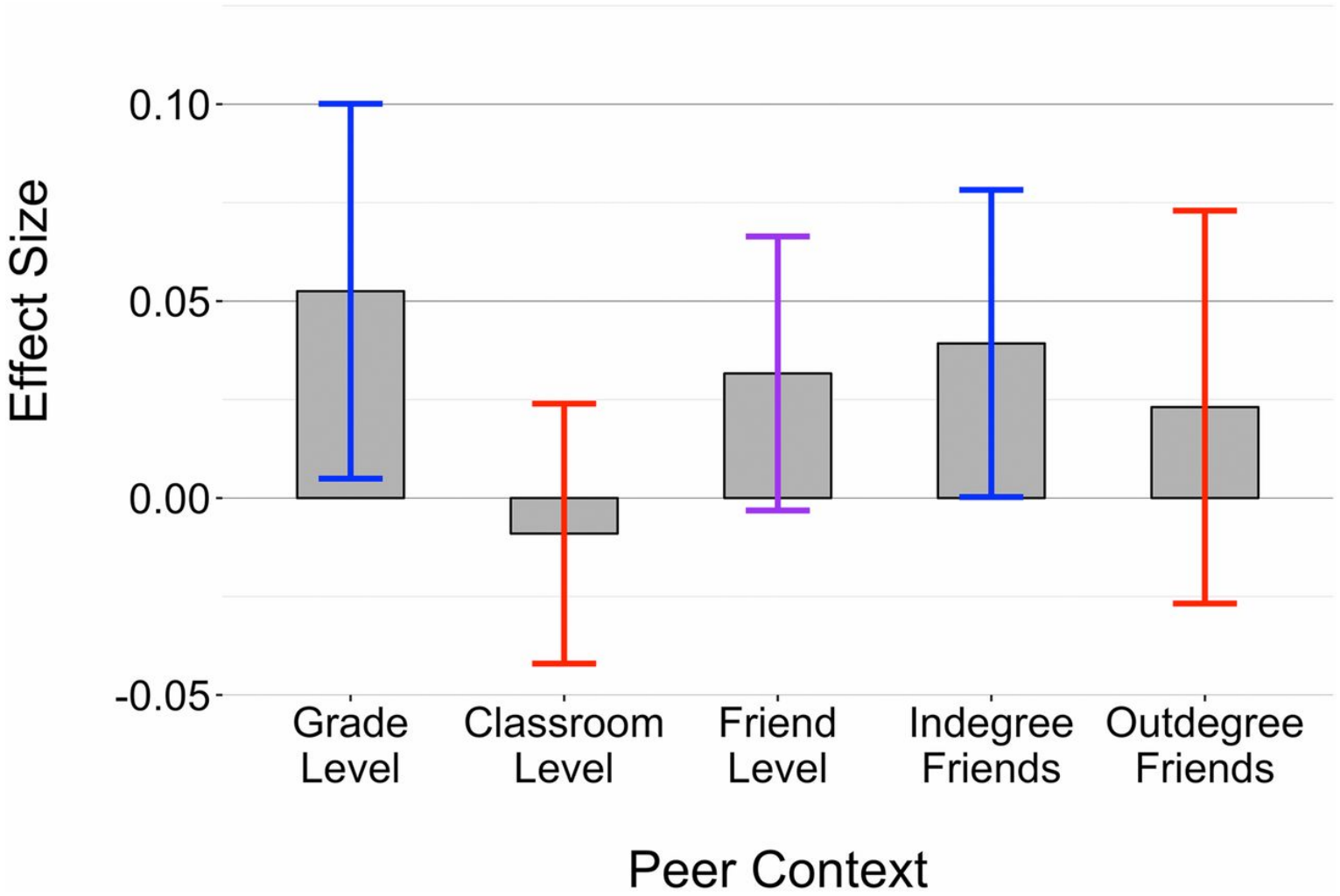
- **Implications**: if there is endogenous sorting, randomly allocating pupils to classes might improve girls' outcomes (but only in the short term)

Genetic peer effects on smoking behaviour

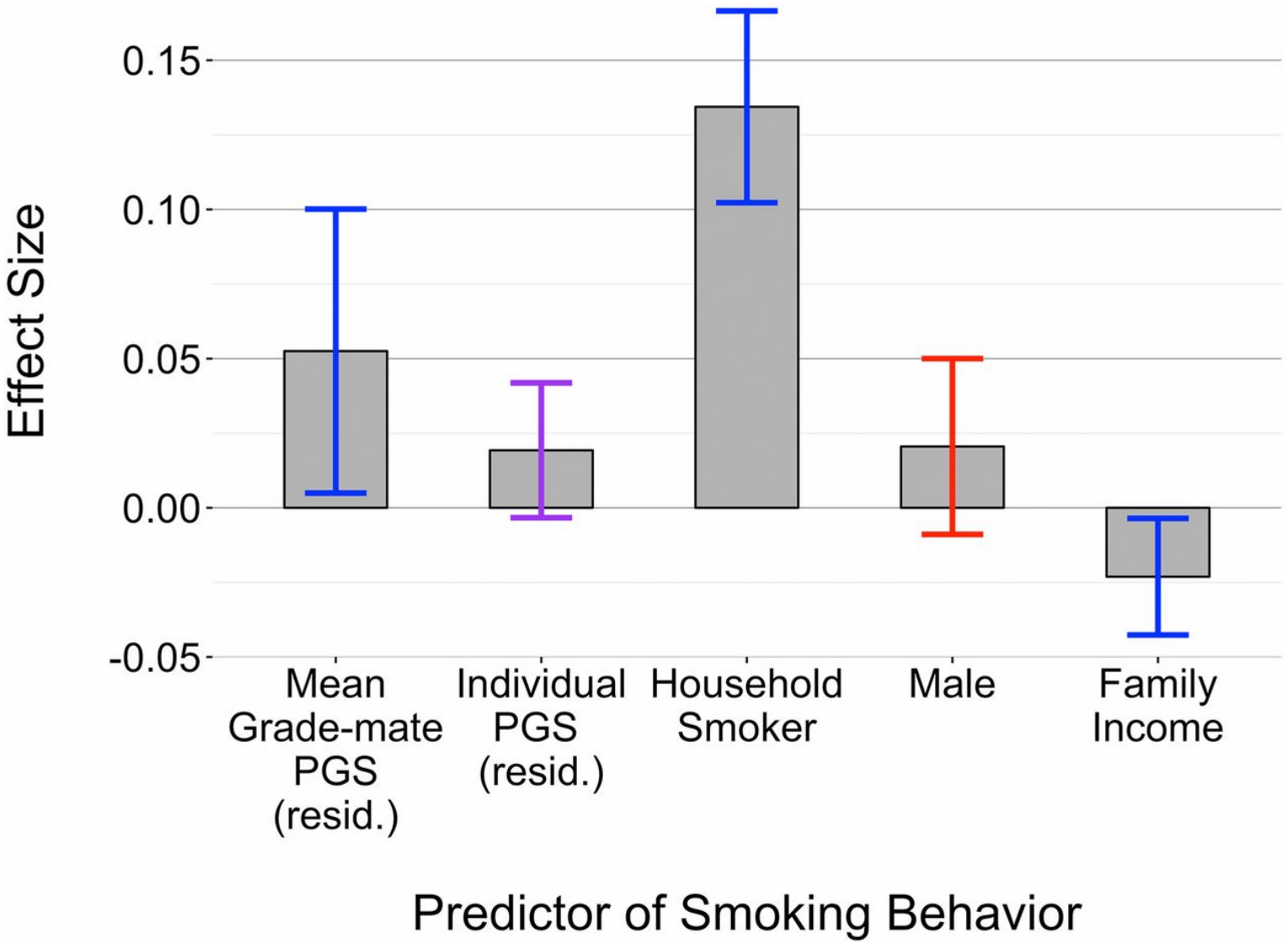
Sotoudeh, R., Harris, K. M., & Conley, D. (2019). Effects of the peer metagenomic environment on smoking behavior. *Proceedings of the National Academy of Sciences*, 116(33), 16302-16307.

- **Dataset**: Add Health
- **PGI for smoking** (No. cigarettes per day)
  - GWAS: Tobacco and Genetics Consortium, Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nat. Genet. 42, 441–447 (2010).
- **Outcome**: No. cigarettes per day (0.15 correlation with PGI)
- **Peer group**: Same grade in school (grade-mates)

Peers smoking PGI and own smoking



Comparison with other predictors of own smoking



Genetic peer effects on educational attainment

Golin, M., & Menta, G. (2025). *Genes, peers and grades*. Mimeo.

Work in progress!

- **Dataset**: the Avon Longitudinal Study of Parents and Children (ALSPAC), $\approx$ 10,000 genotyped children
- **PGI for Educational Attainment**:
  - Based on EA4 GWAS (Okbay *et al.*, 2022). Formula
- **Outcomes**:
  - Key-Stage test scores at age 11
  - Follow up at ages 13 and 16
- **Peer group**: Same grade in school (grade-mates)
- TBI: controlling for parental PGI

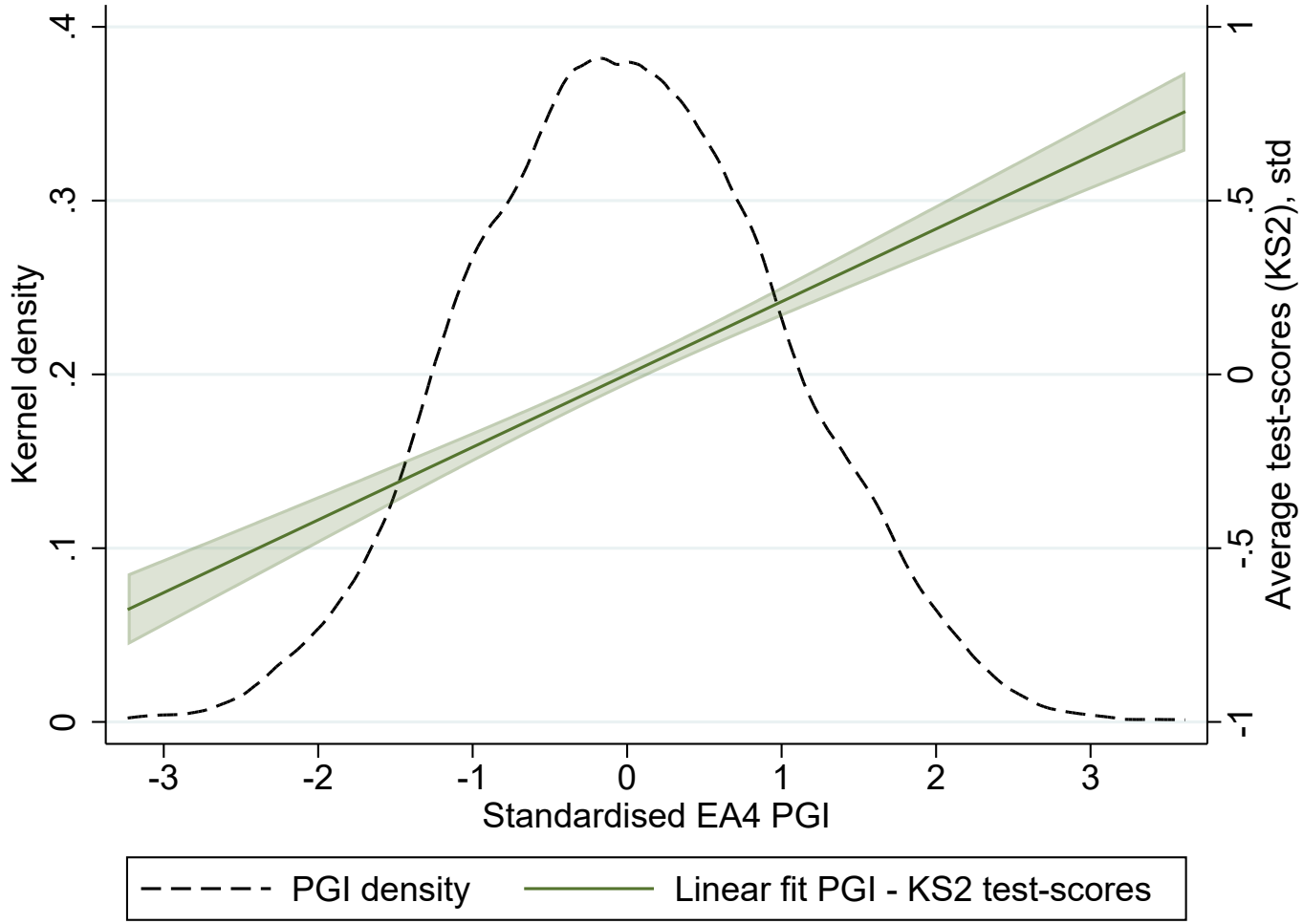
Genetic peer effects on educational attainment

Golin, M., & Menta, G. (2025). *Genes, peers and grades*. Mimeo.

Work in progress!

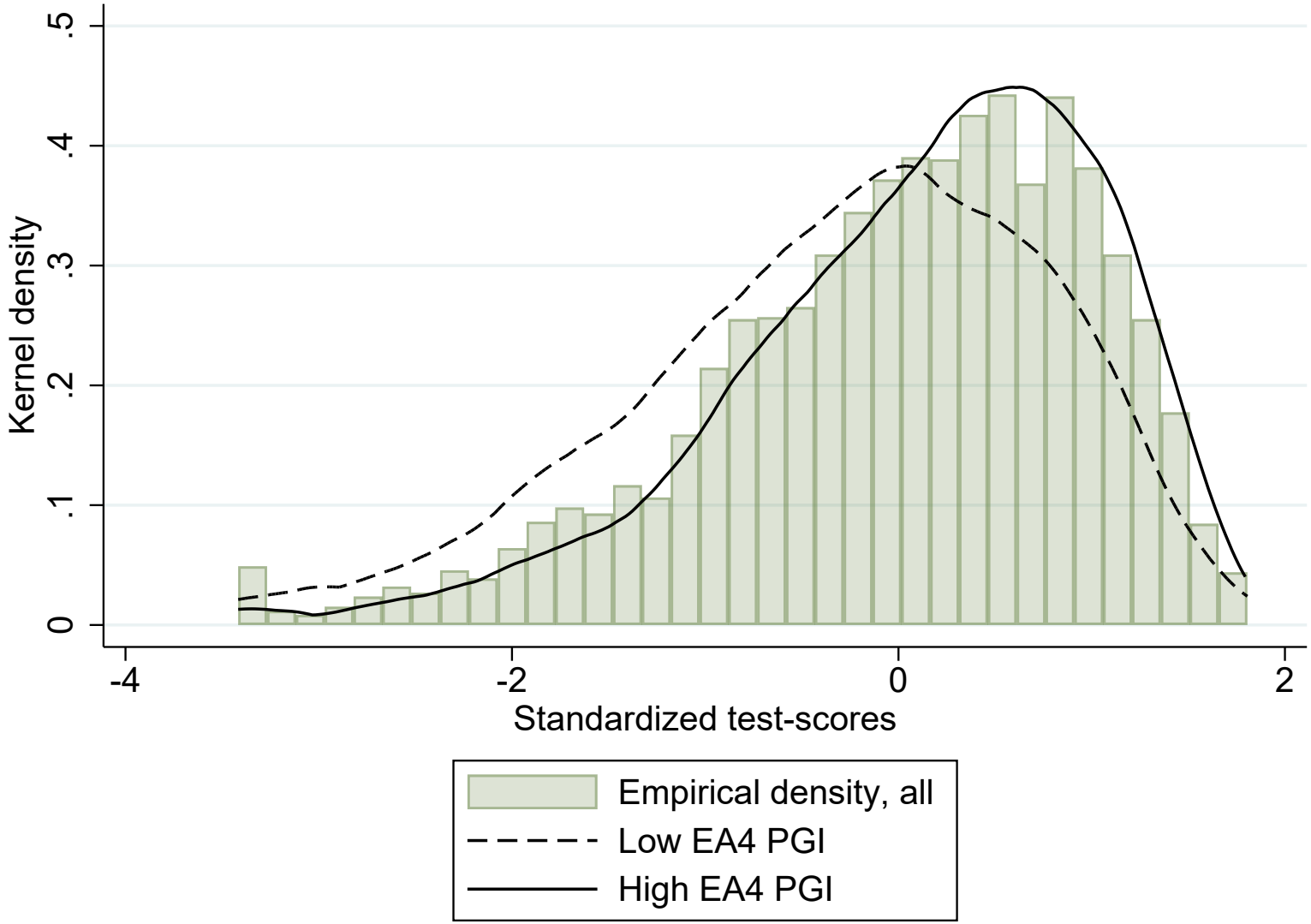
- **Dataset**: the Avon Longitudinal Study of Parents and Children (ALSPAC), $\approx$ 10,000 genotyped children
- **PGI for Educational Attainment**:
  - Based on EA4 GWAS (Okbay *et al.*, 2022). ◂ Formula
- **Outcomes**:
  - Key-Stage test scores at age 11
  - Follow up at ages 13 and 16
- **Peer group**: Same grade in school (grade-mates)
- TBI: controlling for parental PGI

Distribution of EA4 PGI and school achievement



Slope 0.21, p<0.000

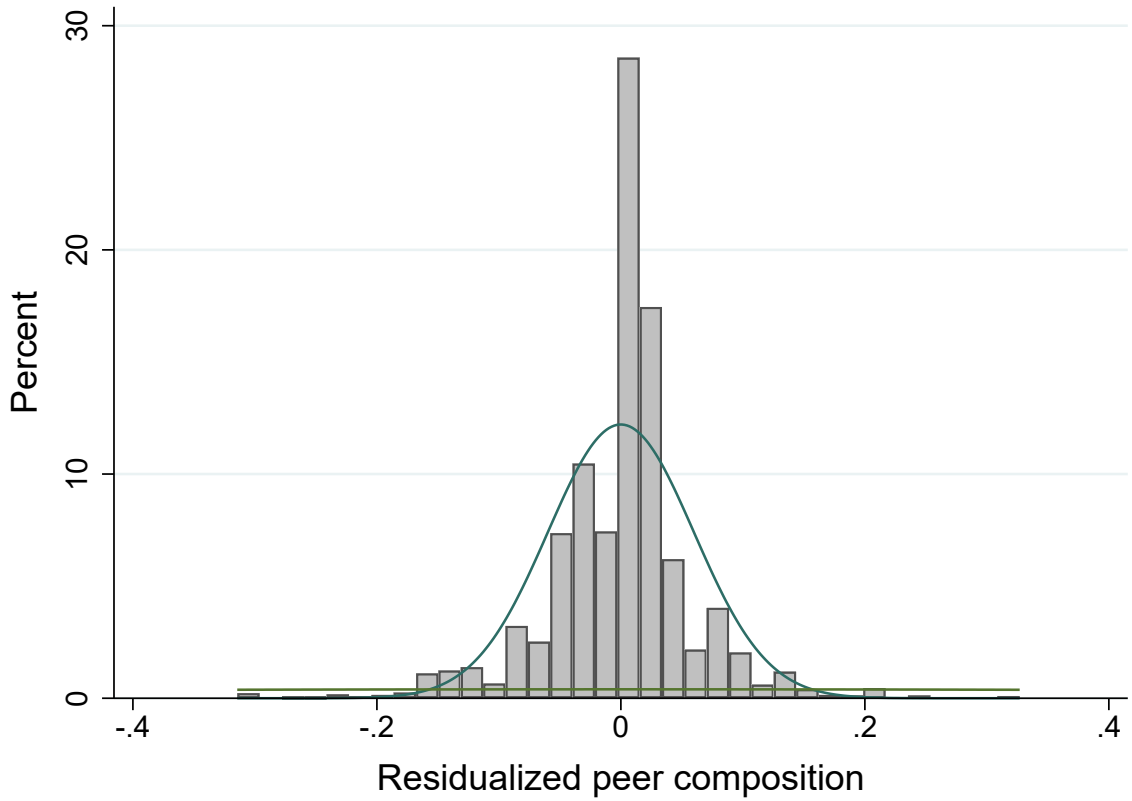
Distribution of KS2 test scores, by PGI status



Identification

We exploit idiosyncratic changes in the share of low-PGI peers, after school and school-year fixed-effects.

Figure: Share of low PGI peers - Residualized



Identification

Let $PGI_i^L = \mathbb{1}_{PGI_i \leq Q1}$, where $Q1$ is the 25^{th} percentile of the PGI distribution. Then, for child i in cohort of size K:

$$KS2_i = \beta_0 + \beta_1 PGI_i^L + \beta_2 \frac{\sum_{k \neq -i} PGI_k^L}{(K-1)} + \beta_3 X_i + \beta_4 \overline{X}_{-i} + \gamma_s + \lambda_t + \eta + \epsilon_i$$

- $KS2_i$ is the Key-Stage 2 average test-score of child i
- X_i includes mother's age at birth, mother' education, father's social class, and dummies for the child's gender, birth month, birth year, and birth order.
- $\overline{X}_{-i}$ leave-one-out average of all controls, within cohort
- γ_s and λ_t are school and school-year fixed effects
- η first 10 ancestry-informative genetic principal components
- s.e. clustered at the cohort level

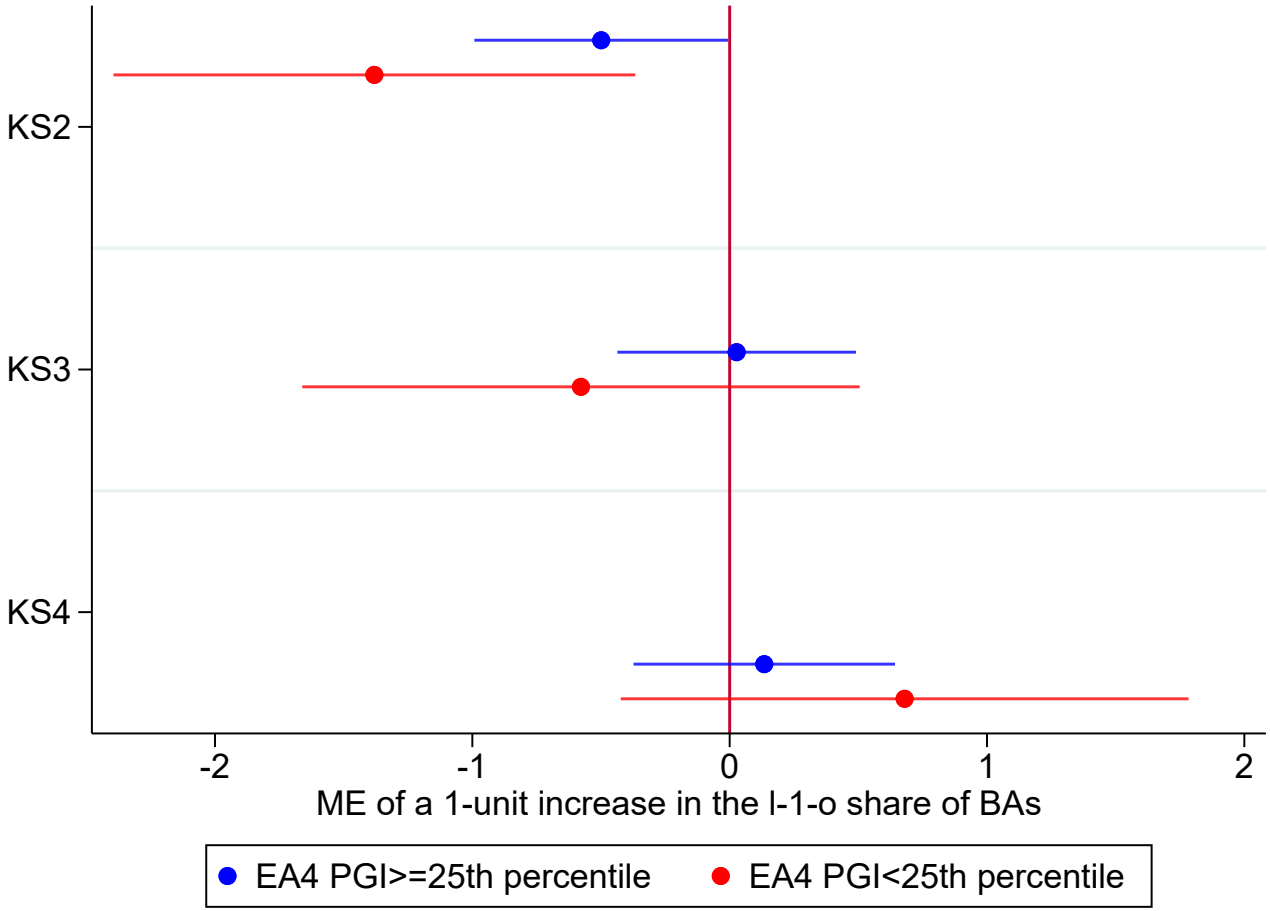
Main results

Figure: Effect of the share of low PGI peers on KS2 test scores

	(1)	(2)	(3)	(4)	(5)
	Panel 1: KS2				
Share of other peers with low education PGI (bottom 25 %ile)	-1.14***	-0.61**	-0.55**	-0.70***	-1.59***
	(0.23)	(0.26)	(0.25)	(0.26)	(0.51)
Low education PGI (bottom 25 %ile)	-0.39***	-0.36***	-0.29***	-0.29***	-0.35***
	(0.04)	(0.04)	(0.03)	(0.03)	(0.04)
Share of other school peers with low educ PGI					-1.78*
					(0.98)
School and school-year FE	No	Yes	Yes	Yes	Yes
Individual controls	No	No	Yes	Yes	Yes
Cohort controls	No	No	No	Yes	Yes
Observations	3,966	3,965	3,965	3,965	3,944
School cohorts	233	232	232	232	230
Adjusted R-squared	0.05	0.16	0.26	0.27	0.27

Main results

Figure: Effect of the share of low PGI peers on test scores, by PGI status



Some of the next steps

- **On the genetics side**:
  - Use Bayesian methods to model LD (e.g. LDpred)
  - Use available trios and impute missing paternal genotypes to control for parental PGIs
  - Correct for measurement error in the PGI (e.g. Becker at al., 2021) to scale estimates by their attenuation factor
- **On the economics side**:
  - Look at the top end of the EA4 PGI distribution ('bright stars')
  - Class level analysis in KS2 (class identifiers available)
  - Explore mechanisms

Some of the next steps

- **On the genetics side**:
  - Use Bayesian methods to model LD (e.g. LDpred)
  - Use available trios and impute missing paternal genotypes to control for parental PGIs
  - Correct for measurement error in the PGI (e.g. Becker at al., 2021) to scale estimates by their attenuation factor
- **On the economics side**:
  - Look at the top end of the EA4 PGI distribution ('bright stars')
  - Class level analysis in KS2 (class identifiers available)
  - Explore mechanisms

Conclusion

- The genetic makeup of peers matters for own health and educational outcomes
- Sanz-de-Galdeano & Terskaya (2023): *parental investments compensate for genetic inequalities* among siblings
- **Less (genetic) segregation** into school-grades might:
  - Reduce girls' risk of obesity (Brunello et al., 2020)
  - Improve educational attainment (Golin & Menta, 2025)
- Need to take into account **behavioral responses** to peer composition: effect of interventions can sometime backfire (e.g. Carrell, Sacerdote & West, 2013; Booij, Leuven & Oosterbeek, 2017)

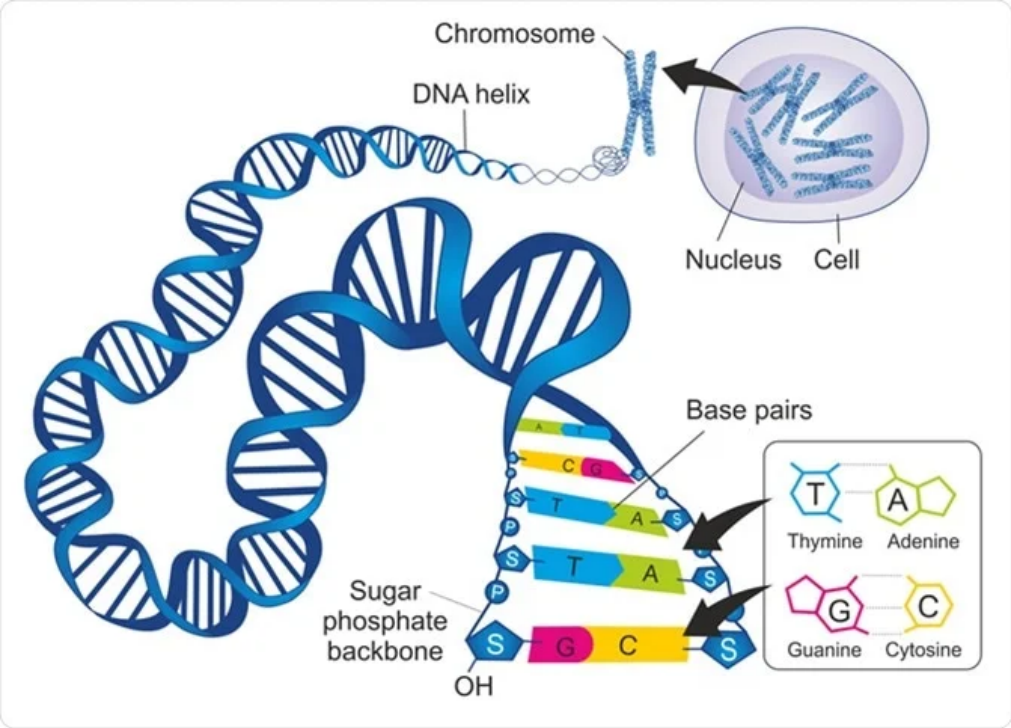
The End

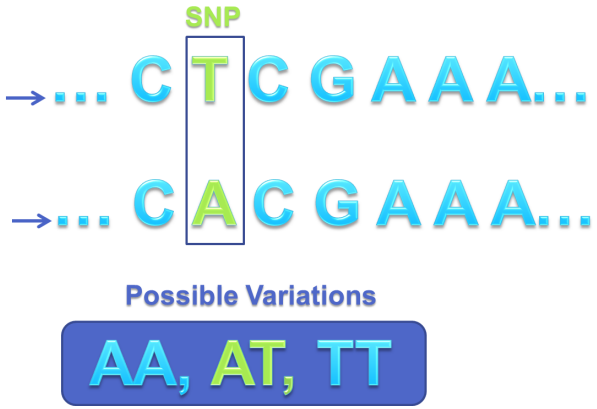
Thank You!

Questions?

giorgia.menta@liser.lu

The structure of genetic data





- 99% of DNA is the same across individual, except for **Single Nucleotide Polymorphisms** (**SNPs**): common variations in human DNA
- **Effect Allele**: allele associated with the trait or disease
  - How many X effect alleles are there in a given SNP? $X \in \{0, 1, 2\}$

The Polygenic Index

Back

The PGI is constructed as a **weighted sum** of all SNPs involved in a trait. We build it using the terminal program PLINK 1.9, which uses the following formula:

$$PGI_j = \frac{\sum_{i=1}^{N} \hat{\beta}_i \times X_{ij}}{P \times M_j}$$

for each individual j.

- $\hat{\beta}_i$: effect size of SNP i on the trait of interest (retrieved from published GWAS)
- X_{ij}: no. of effect alleles in individual j for SNP i; $X_{ij} \in \{0, 1, 2\}$
- P: ploidy of the sample (generally 2 for humans)
- M_j: number of non-missing SNPs observed in individual j