GWAS for social science outcomes

Philipp Koellinger

University of Amsterdam &
Erasmus University Rotterdam

University of Chicago
18 Apr 2014

GWAS meta-analyses in the social science

- Cognitive performance
  - Benyamin et al. (2013): $N = 17,989$; null-result
- Personality
  - De Moor et al. (2012): $N = 20,669$; null-result
- Self-employment
  - Van der Loos et al. (2013): $N = 53,838$; null-result
- Will this ever work???
  - Effect sizes *too* small?
  - Effects not linear?
  - G × E?
  - Twin studies wrong?

GWAS educational attainment
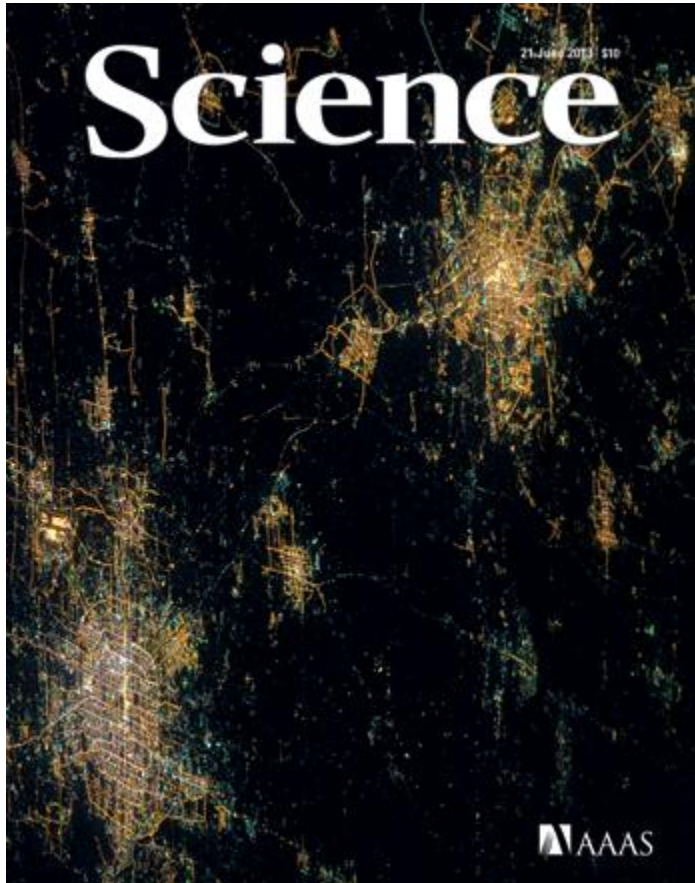
GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment

Rietveld et al., *Science*, 340, 1467-1471, 2013

GWAS on educational attainment

- 42 + 12 = 54 cohorts participated ($N \sim$ 125,000)
- Two measures for educational attainment
  - Years of schooling (*EduYears*)
  - College degree yes/no (*College*)
- Strict quality controls
  - Data
    - MAF, call rate, imputation quality etc.
  - Population stratification
    - First 4 principle components from cohort-specific genetic relatedness matrix
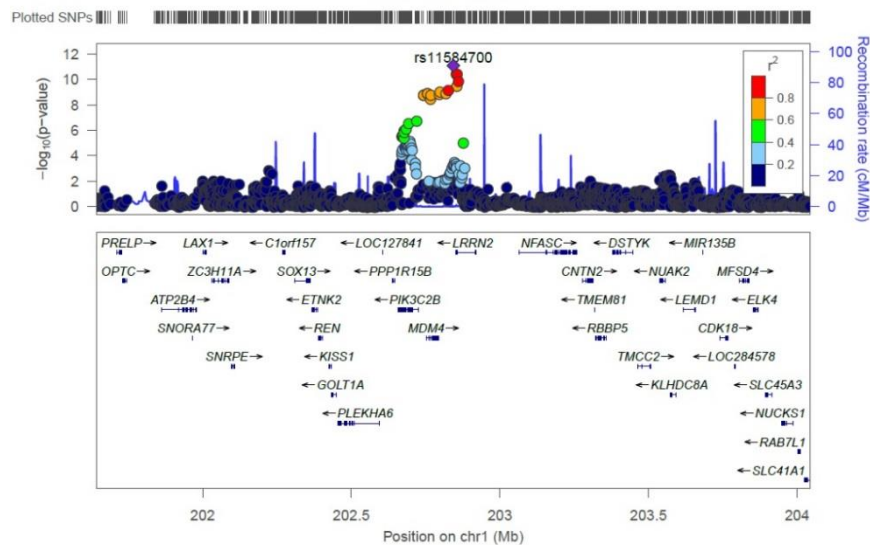   - Genomic control
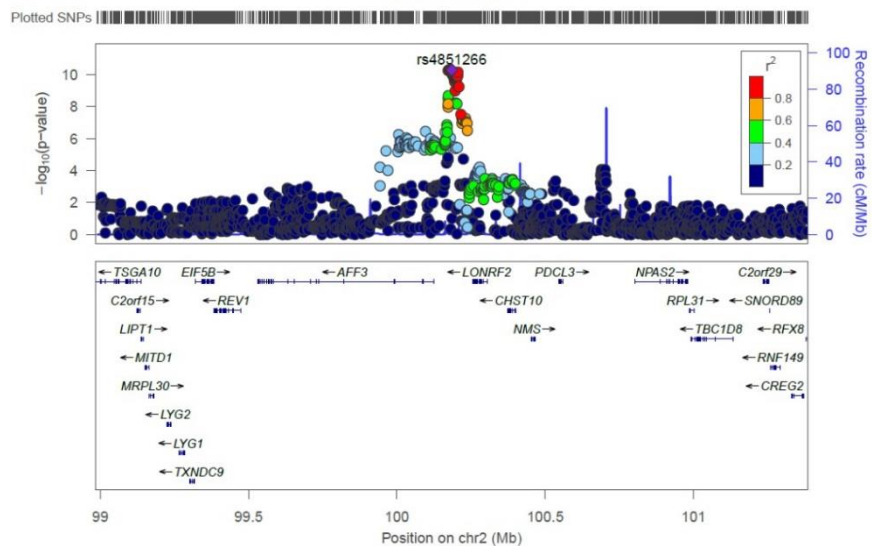- Many follow-up analyses

Three replicated SNPs
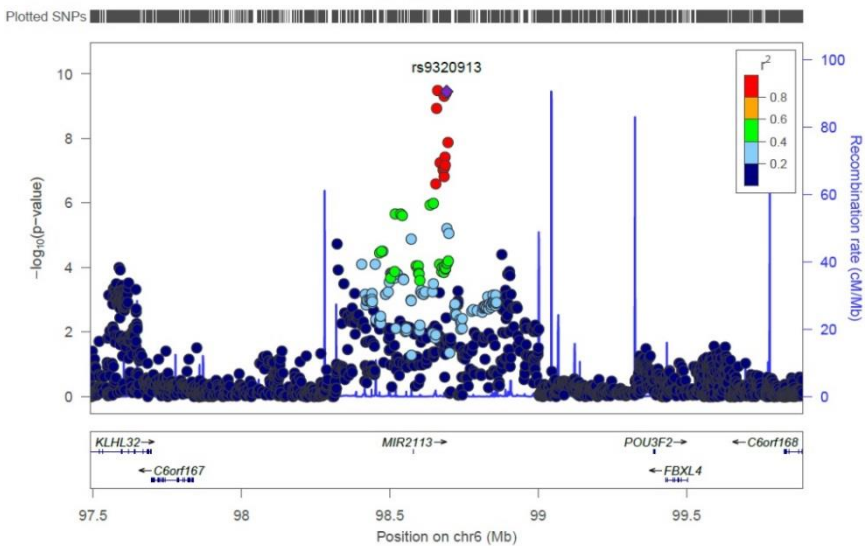
EduYears

College

Replicated loci from discovery stage

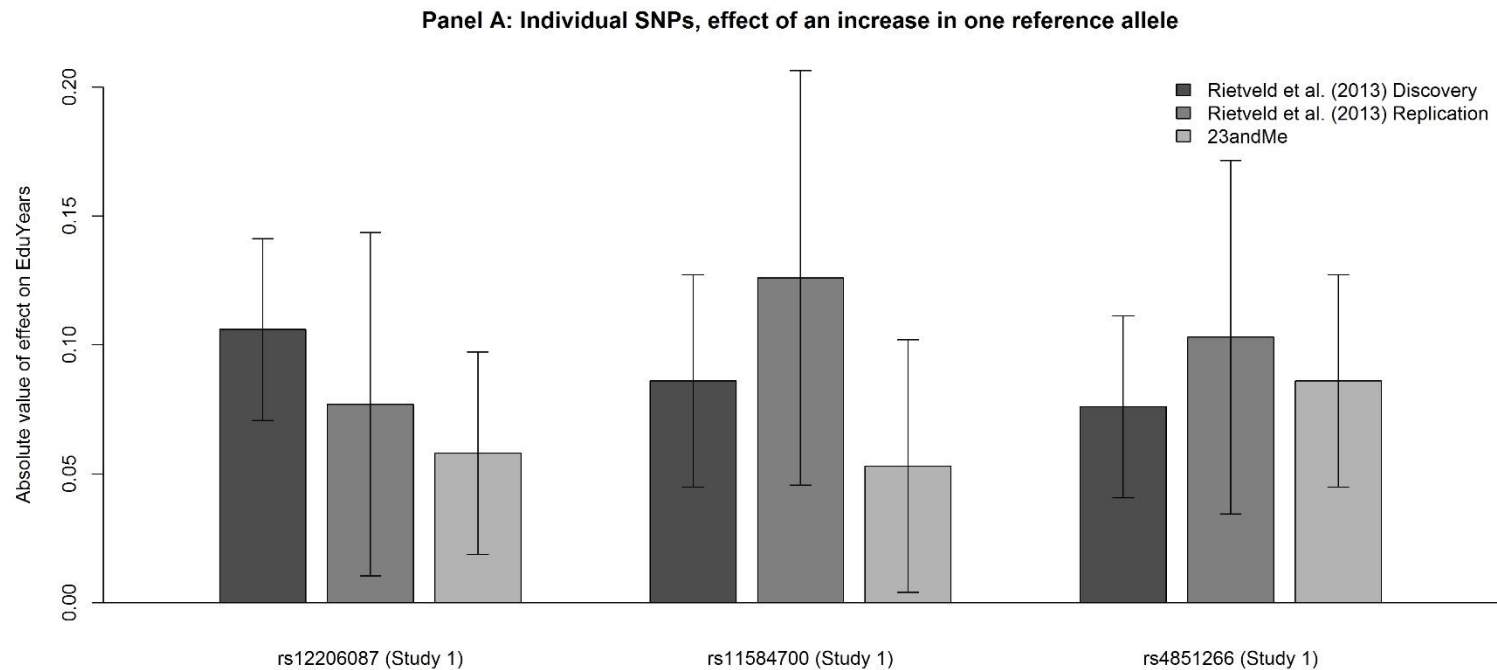|		Discovery		Replication		Combined	
	SNP	*Beta*	*p*-value	*Beta*	*p*-value	*Beta*	*p*-value
College	rs11584700	-0.101	2.07×10^{-09}	-0.101	0.001	-0.101	8.24×10^{-12}
	rs4851266	0.064	2.20×10^{-09}	0.075	0.007	0.066	5.33×10^{-11}
EduYears	rs9320913	0.076	4.19×10^{-09}	0.062	0.024	0.076	3.50×10^{-10}

- Largest effect sizes of replicated SNPs:
  - Max R^2 = 0.022% for *EduYears,* ~2 months difference between the two homozygotes
  - Max *Odds* = 0.904 for *College*, ~5.4% difference in chance to complete college between the two homozygotes

Polygenic scores for educational attainment

- R^2 increases with number of SNPs included in the score
- Currently ($N \approx 120{,}000$) $R^2 \approx 2\%$ for EA ($p < 10^{-28}$)
- EA score has $R^2 \approx 2.5\%$ for cognitive function ($p < 10^{-9}$)
- Projections for accuracy of polygenic scores for EA:
    - With $N \approx 500{,}000 \rightarrow R^2 \approx 12\%$
    - With $N \approx 1{,}000{,}000 \rightarrow R^2 \approx 15\%$
    - With $N \rightarrow \infty \rightarrow R^2 \approx 22\%$

Is this real?

- Genome-wide significant SNPs replicate in additional sample that used very stringent controls for population stratification
  - 23andMe ($N = 34,428$), controlling for 25 PCs

Panel A: Individual SNPs, effect of an increase in one reference allele

Legend:
- Rietveld et al. (2013) Discovery
- Rietveld et al. (2013) Replication
- 23andMe

Y-axis: Absolute value of effect on EduYears (0.00 to 0.20)

X-axis categories: rs12206087 (Study 1), rs11584700 (Study 1), rs4851266 (Study 1)

Note: Bars are 95% confidence intervals

Is this real?

- Polygenic scores replicate in mixed linear models that allow the correlation between individuals' educational attainment to be linearly increasing in their genetic relatedness
  - QIMR ($N = 3{,}544$) and STR ($N = 6{,}770$)
  - Almost identical R^2 and p-values as in Rietveld et al. (2013)
- Polygenic scores replicate in within-family tests
  - STR (2,774 DZ twins), QIMR (572 full-sib pairs), FHS (395 families with two or more full sibs)

→ Extremely unlikely that association results are driven by population stratification

EA as a proxy-phenotype for cognitive performance

- GWAS on EA with $N = 100,000$ has more power to detect SNPs associated with cog. performance (CP) than GWAS on cognitive performance with $N = 25,000$.
  - Due to much larger sample size and moderate genetic correlation ($r \approx 0.7$, Trzaskowski et al. 2013).

EA as a proxy-phenotype for cognitive performance

- Our 1st stage: GWAS on educational attainment ($N =$ 106,736).

  - The Education Sample (and analysis plan) same as Rietveld et al. (2013), except excluding datasets with high-quality cognitive performance data.

  - Retain as "education-associated SNPs" all SNPs with p-value < 10^{-5} for association with College or EduYears, 69 in total.

- Our 2nd stage: Test the 69 "educated-associated SNPs" for association at $p < .05/69$ with cognitive performance ($N = 24,189$).

  - The Cognitive Performance Sample is larger than the largest GWAS sample size to date (indeed, it is a superset).

 - But still would be too small without the proxy-phenotype approach.

Results
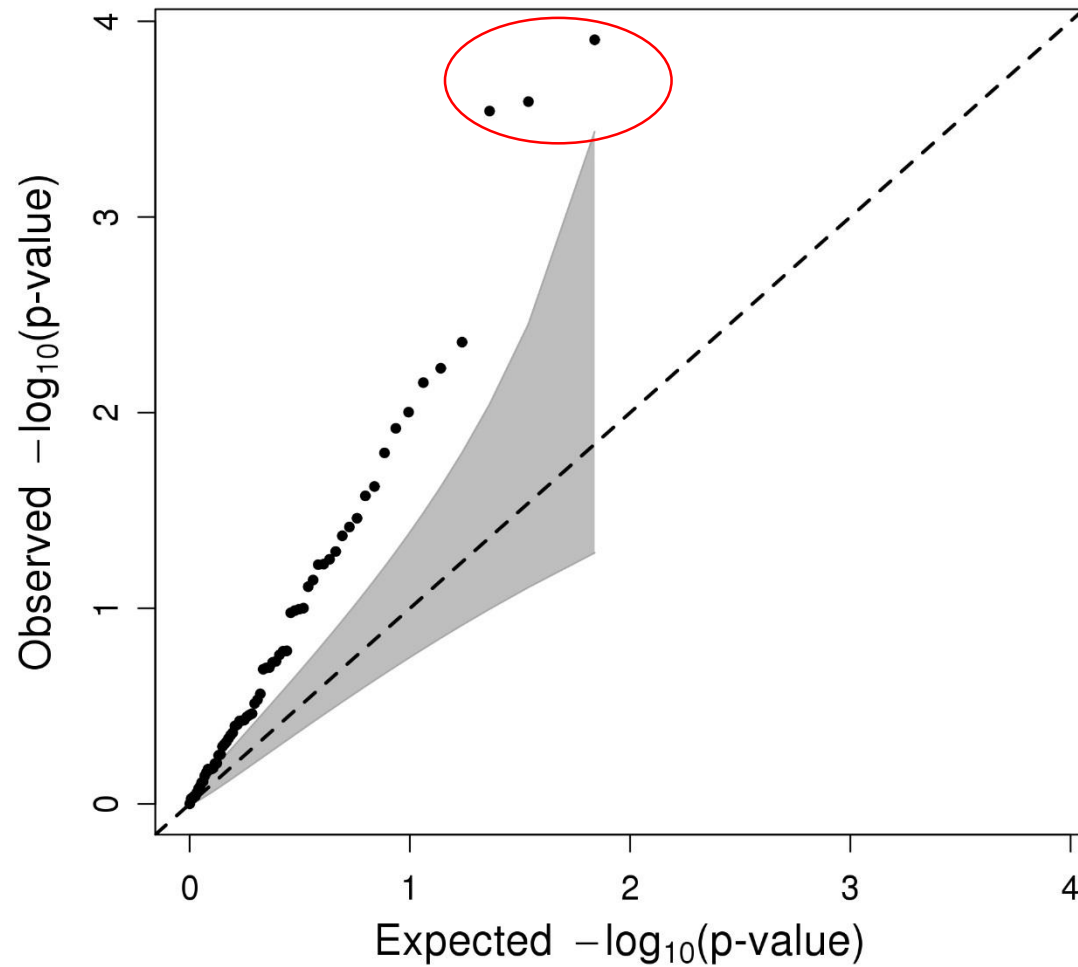
Note: Shaded area is 95% confidence interval around the null hypothesis

Conclusion

- Social-scientific outcomes are most likely highly polygenic traits
- GWAS can work
  - But large sample sizes and careful study designs are needed
- Ongoing GWAS projects of the SSGAC (www.ssgac.org)
  - Educational attainment 2.0
    - $N \approx 200,000$
  - Subjective well-being
    - $N \approx 140,000$
  - Fertility (number of children, age at first birth)
    - $N \approx 130,000$

Results
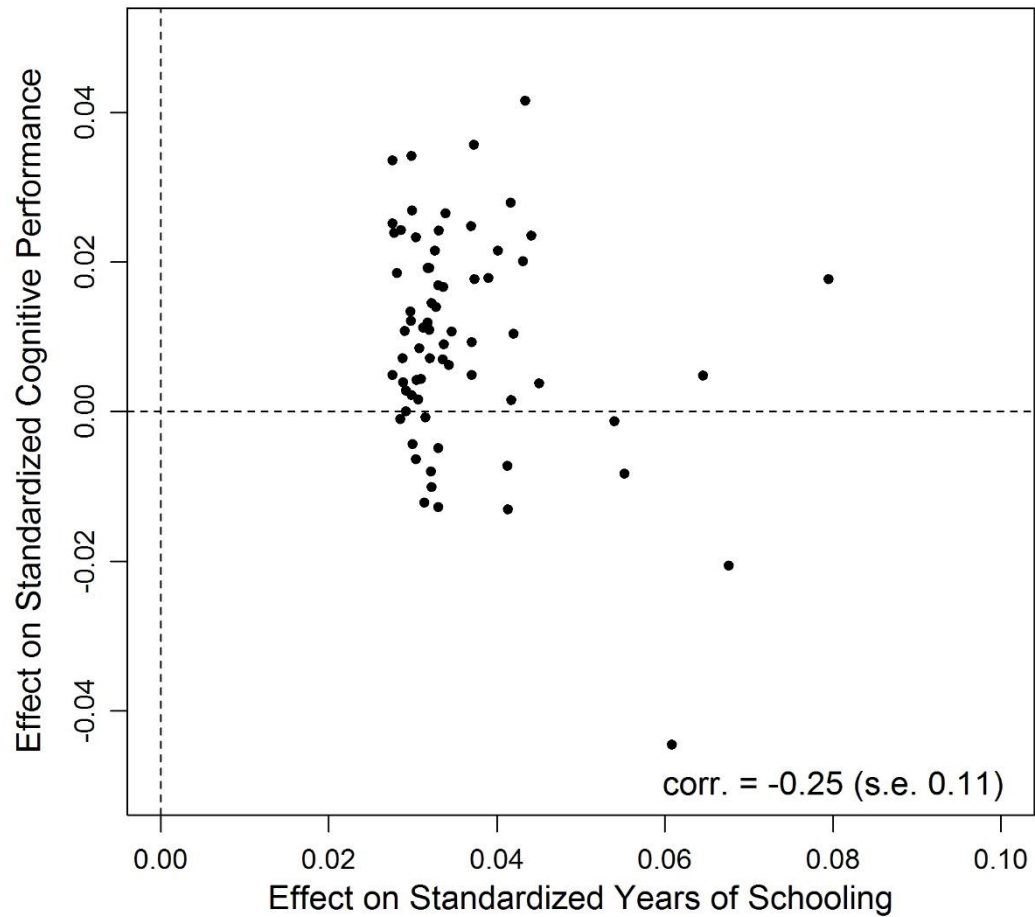
Note: The relationship between standardized coefficients from the first-stage regression of years of schooling on the education-associated SNPs in the Education Sample (x-axis) and standardized coefficients from the second-stage regression of cognitive performance on these SNPs in the Cognitive Performance Sample (y-axis). The reference allele is chosen such that the coefficient on years of schooling is positive.
Each point represents one of the 69 education-associated SNPs.

Polygenic score of 69 EA SNPs and mental health
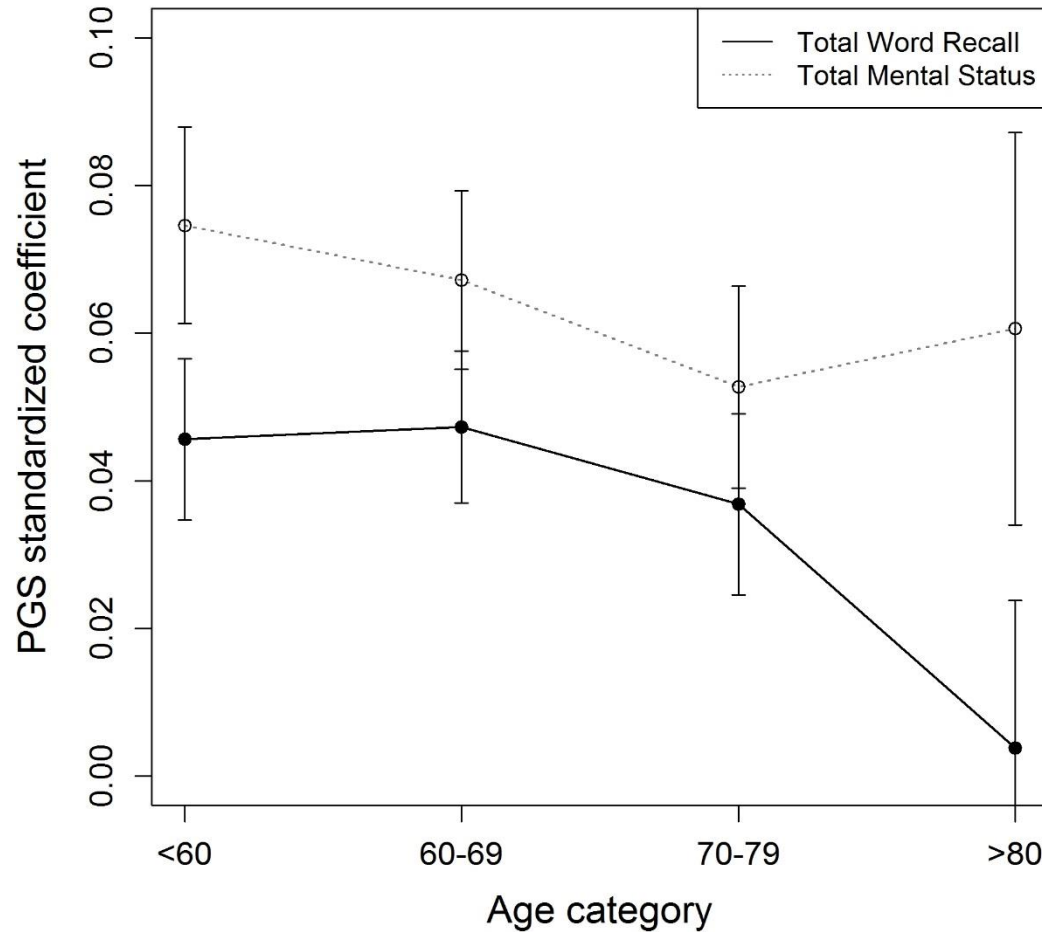
Note: Coefficients from regression of standardized cognitive phenotype (Total Word Recall or Total Mental Status) on standardized polygenic score within age category, controlling for sex and clustering standard errors by individual. Error bars show ±1 standard error.

Biological annotation of EA genes

- EA genes over-expressed in anterior caudate nucleus
- Links to immune function, body size, and inflammation
- Co-expression analyses:
  - Neuronal pathways (*BSN, GBX2, LRRN2, PIKC2B*)
  - Neuron fate (*GBX2*)
  - Synaptic terms (*BSN, LRRN2*)
  - Learning and long-term memory (*BSN*)
  - Muscular contractions and neuron-muscle junctions (*PIK3C2B, IP6K3, ITPR3, TET2*)
- Animal model databases:
  - Anterior hindbrain development in zebrafish and mice (*GBX2*)
  - Central nervous system in mice (*MDM4*)