Results from the Gentrepreneur Consortium

Philipp Koellinger (Economics, Erasmus U Rotterdam)

3rd SSGAC Workshop • 19 May 2012

GWAS on entrepreneurship

- Earliest large-scale GWAS initiative on an economic outcome
  - 17+1 studies
  - Started in 2008
  - Replication stage finished in fall 2011
- Main analysts:
  - Matthijs van der Loos (Rotterdam), Niina Eklund (FIMM, THL, Finland), Niels Rietveld (Rotterdam)
- Two proxies, two research strategies:
  - At least once self-employment (*n* = 59,549 + 3,271)
  - Serial self-employment (*n* = 5,930 + 2,771)
- Why entrepreneurship?
  - It is an economically important phenomenon
  - It is moderately heritable
- Calibrating expectations about genetic effects in economics

Participating studies

- AGES
- ARIC
- ASPS
- ERF*
- FHS
- GHS
- H2000
- HBCS
- KORA S4
- NFBC1966
- NTR
- RS*
- SardiNIA
- SHIP
- STR* **
- THISEAS
- TwinsUK
- YFS

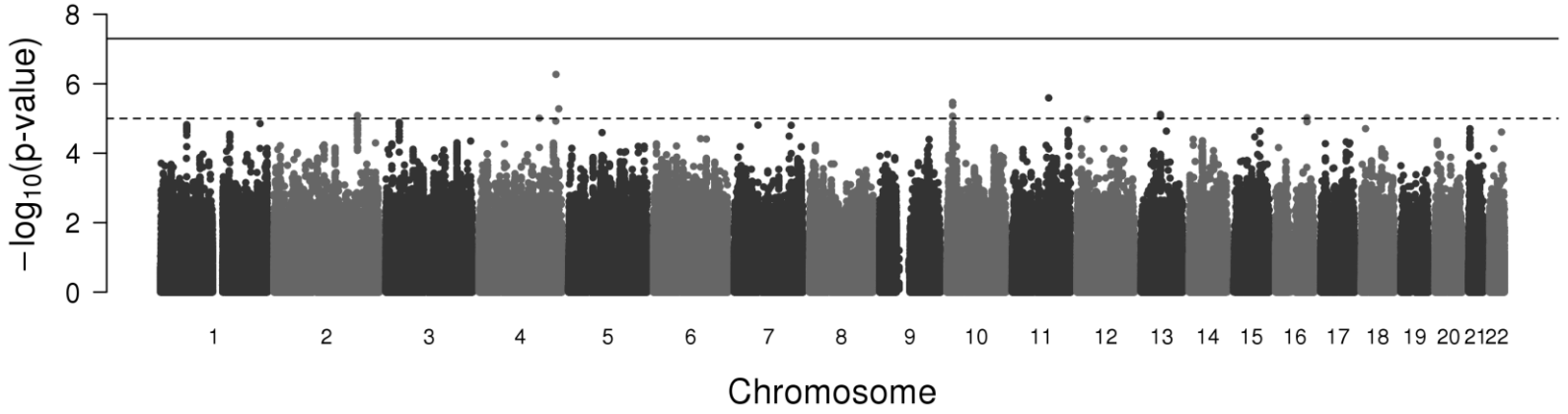
* for studies that also have serial self-employment
** for replication study

Twin study and GREML results

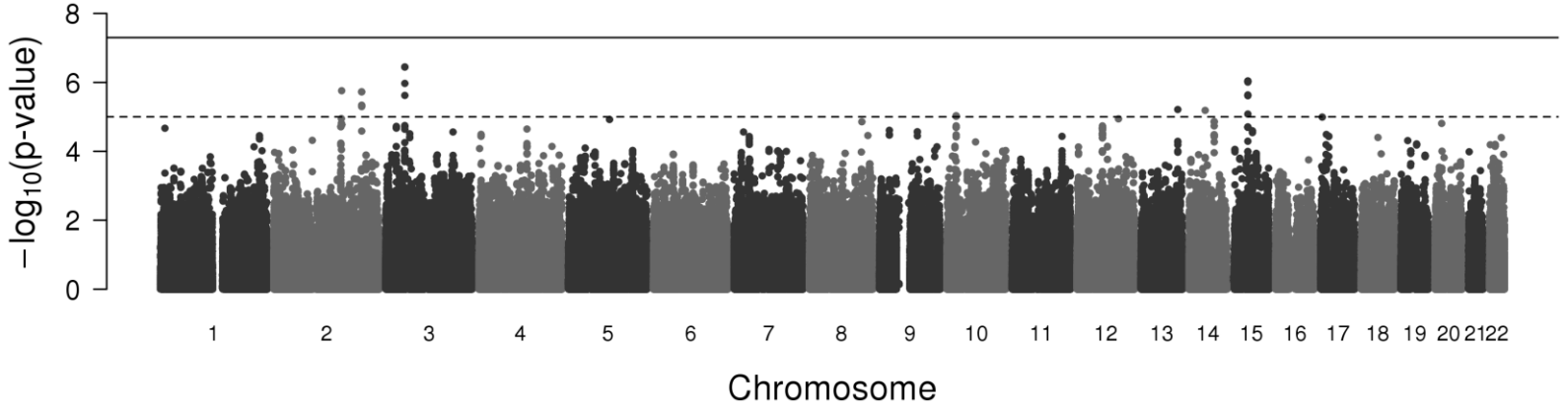
	Twin study			GREML			
Proxy	h^2	**(95% CI)**	***n***	h^2_{SNPs}	**se**	***p***	***n***
Once	0.54	(0.25–0.63)	4,464	0.25	0.14	0.034	6,223
Males	0.67	(0.33–0.76)	1,776	0.26	0.24	0.140	2,986
Females	0.38	(0.00–0.53)	2,688	0.00	0.28	0.498	3,835
Serial	0.61	(0.18–0.82)	3,404	0.17	0.28	0.272	5,672
Males	0.60	(0.13–0.87)	1,224	0.73	0.44	0.047	2,602
Females	0.59	(0.00–0.79)	2,180	0.00	0.70	0.500	3,592

- Twin study estimates for STR
- GREML estimates for RS and STR
- Twin study controls for sex
- GREML controls for sex, birth year, birth year^2, birth year^3, cohort, and the first ten PCs of genotypic data

Manhattan plots discovery stage - pooled

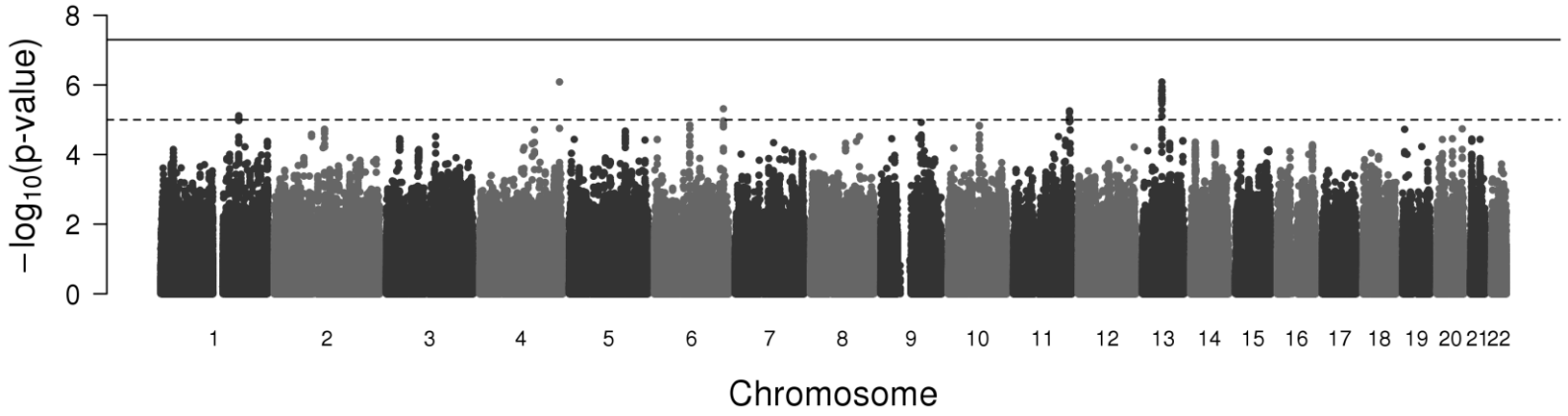


Once self-employment

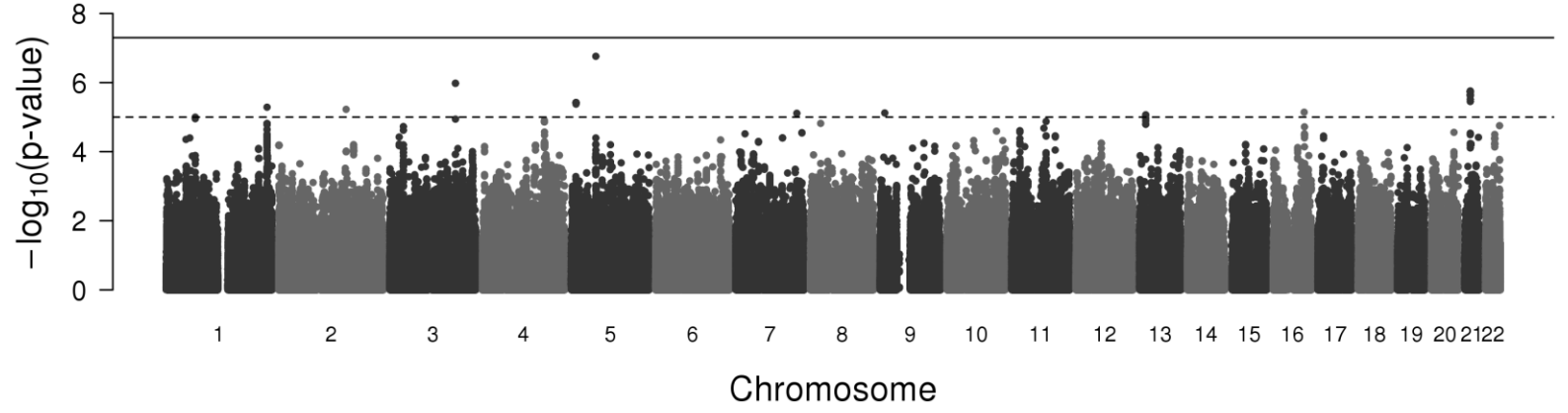


Serial self-employment

Manhattan plots discovery stage - males

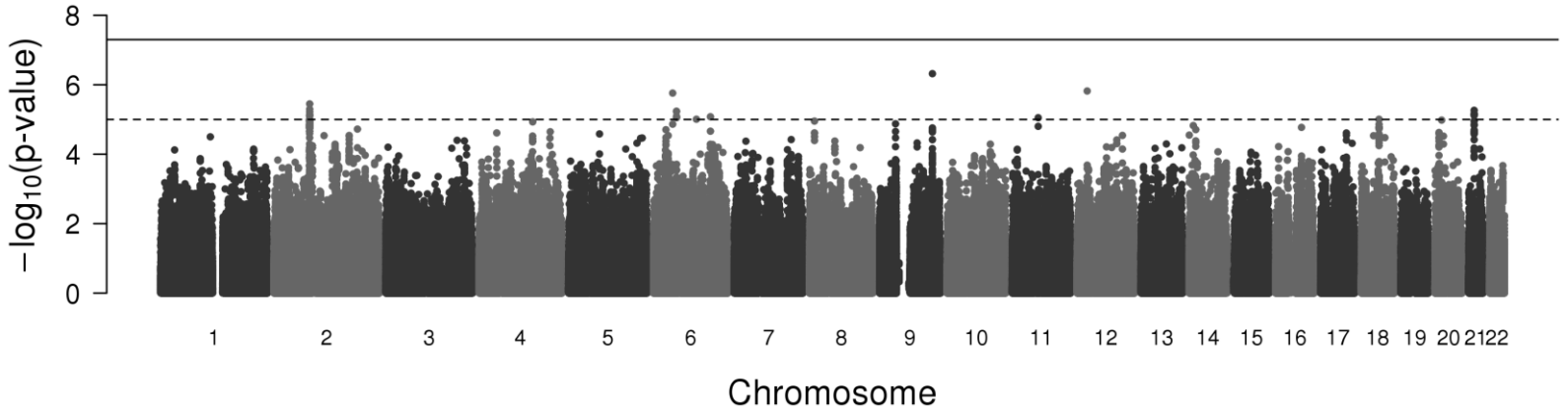


Once self-employment

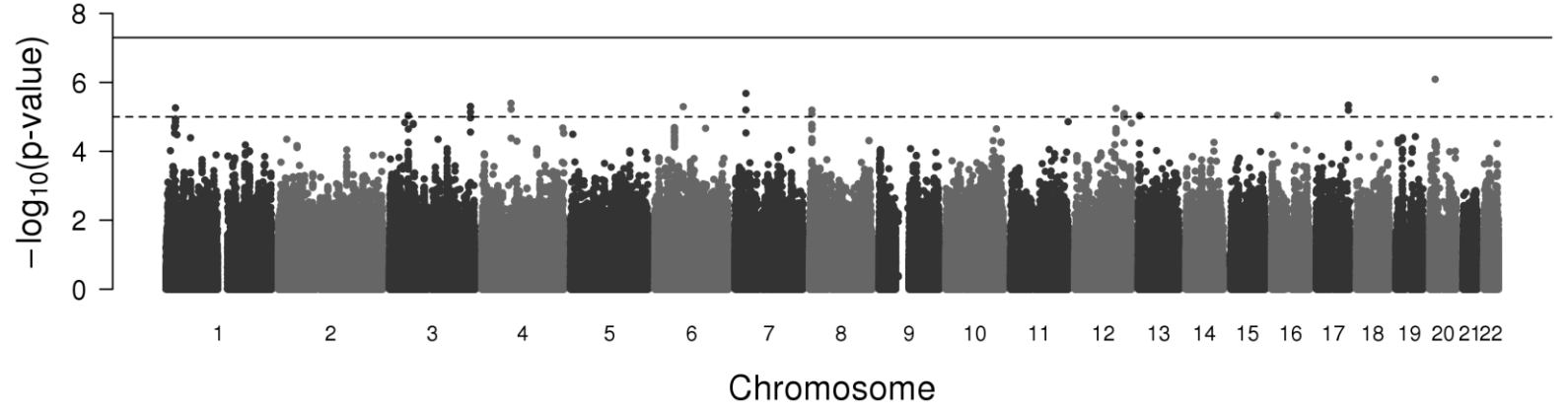


Serial self-employment

Manhattan plots discovery stage females

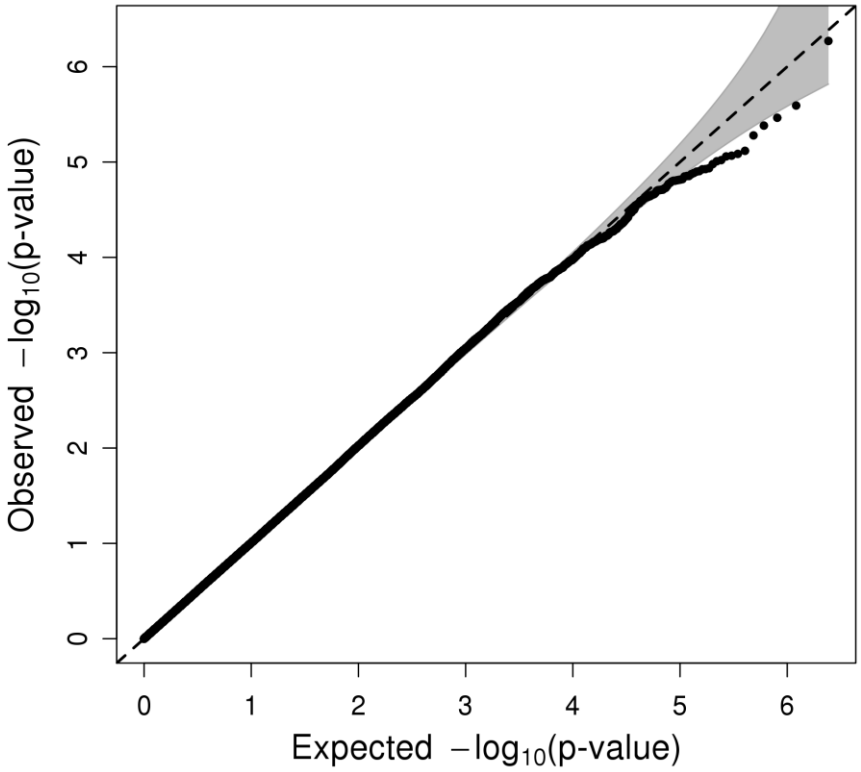


Once self-employment

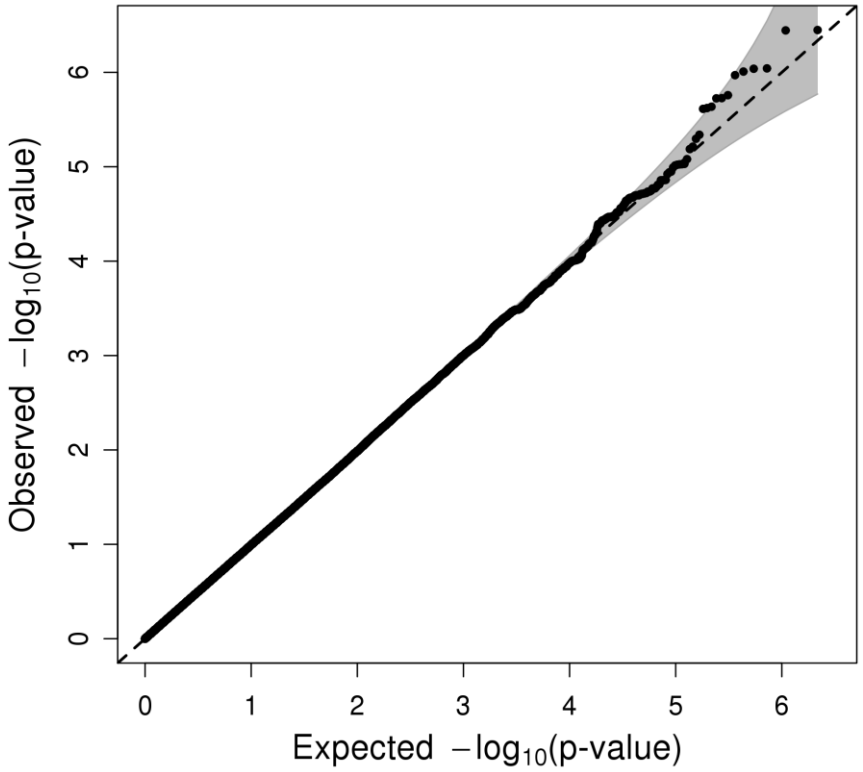


Serial self-employment

Q–Q plots discovery stage – pooled

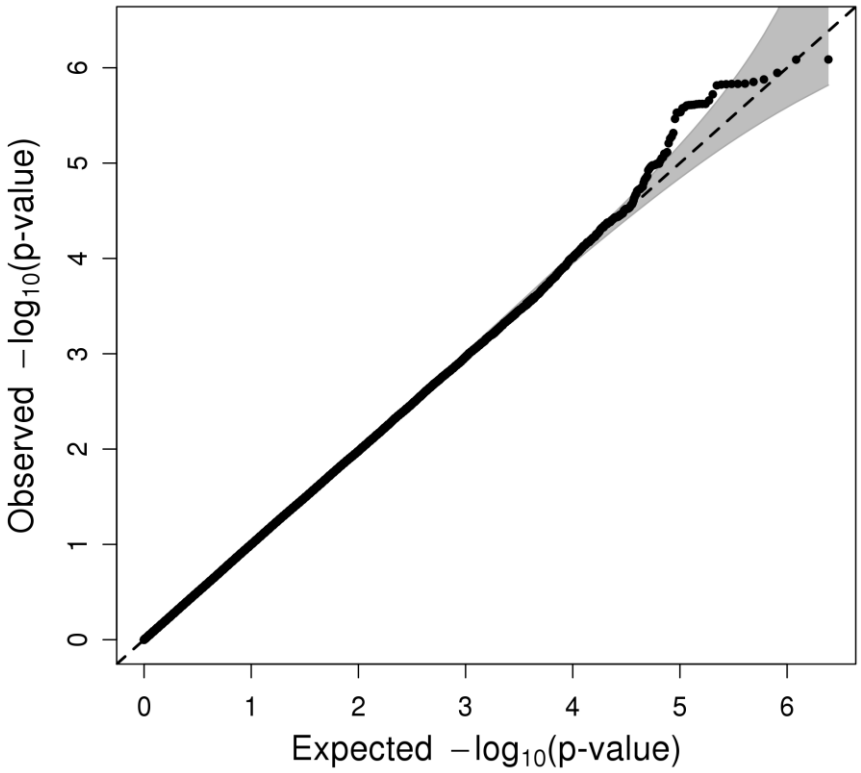


Once self-employment

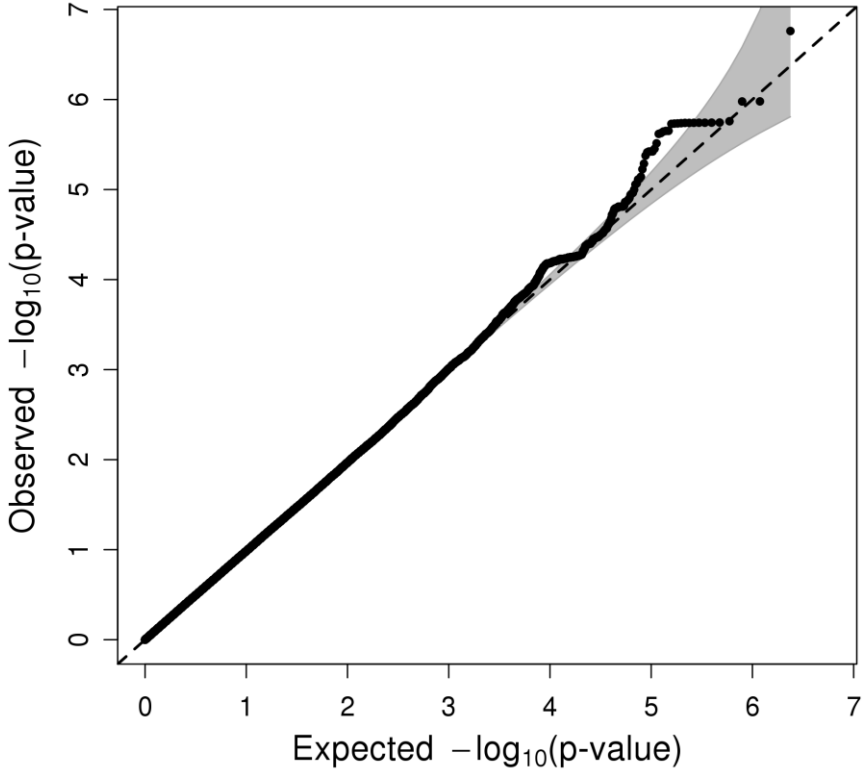


Serial self-employment

Q–Q plots discovery stage - males

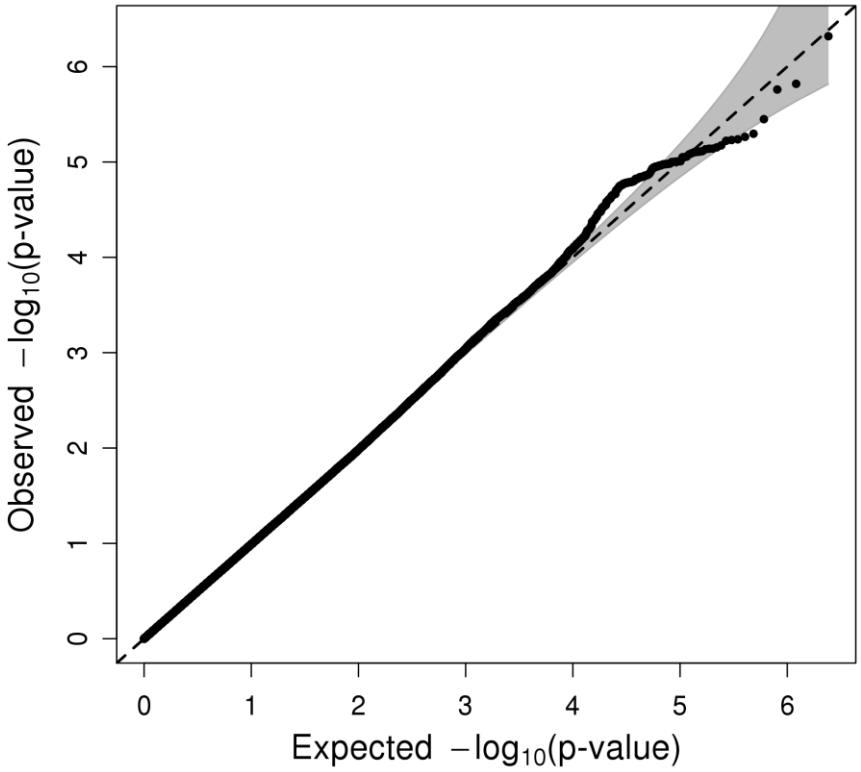


Once self-employment

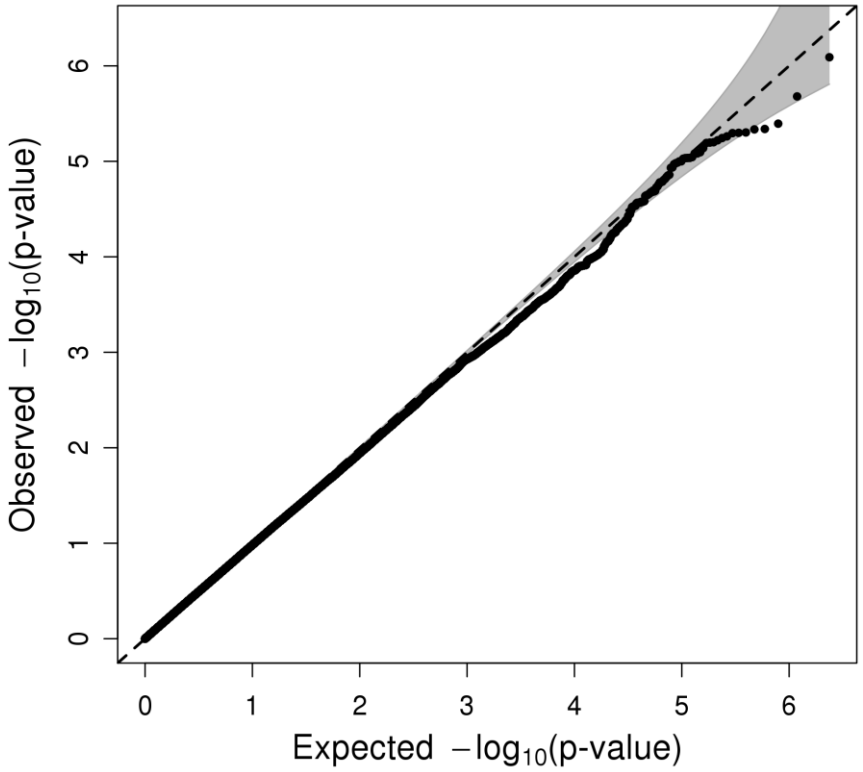


Serial self-employment

Q–Q plots discovery stage females

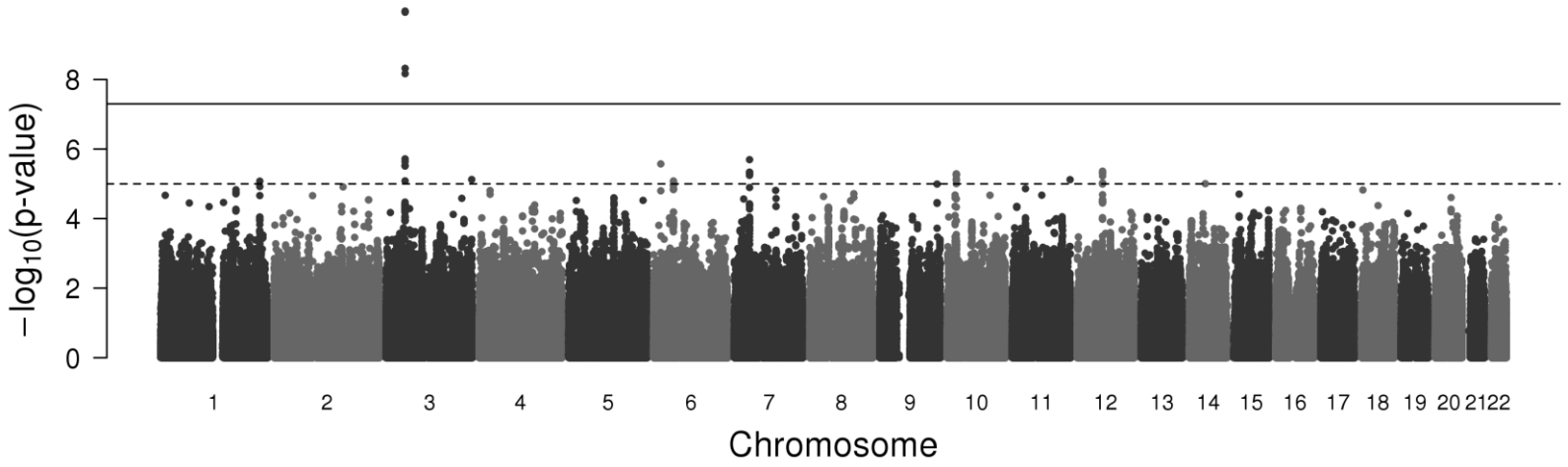


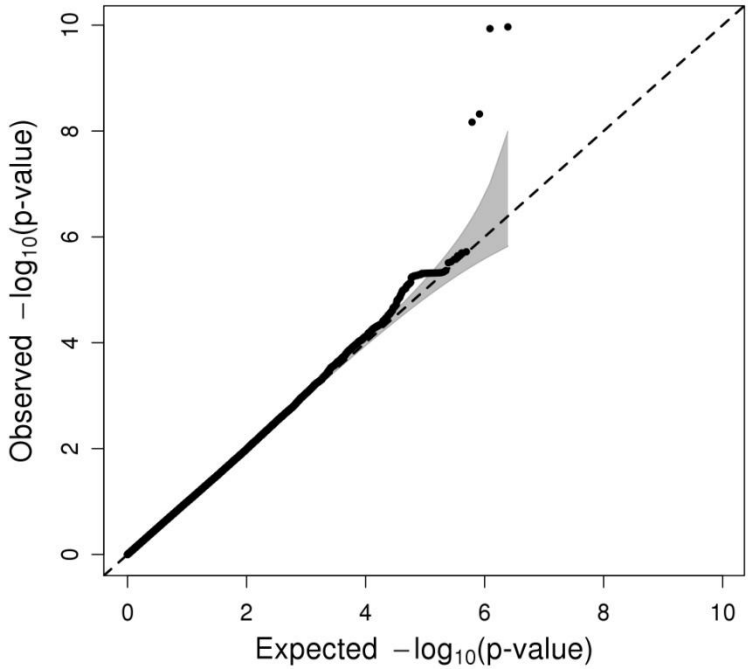
Once self-employment



Serial self-employment

Manhattan plot combined meta-analysis





Serial self-employment

Posterior probability of a true association

- Bayes' rule:

$$\Pr(true \mid significant) = \frac{\Pr(significant \mid true) \times \Pr(true)}{\Pr(significant \mid true) \times \Pr(true) + \Pr(significant \mid false) \times \Pr(false)}$$

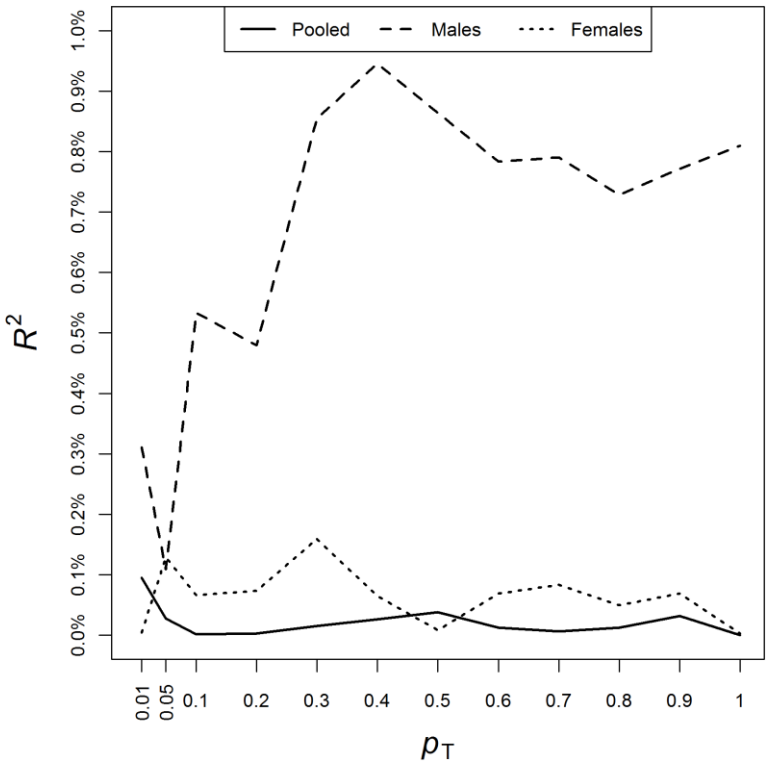
- Pr(*significant* | *false*)
  - Genome-wide significance level: 5×10^{-8}
- Pr(*significant* | *true*)
  - power at 5×10^{-8} in n = 8,701, given OR 1.042: 1.07×10^{-7}
- Pr(*true* | *significant*)
  - > 50% if Pr(*true*) > 32%
  - More realistic prior:
    - 1000 SNPs → prior: 1,000/1,000,000 SNPs = 0.1%
  - Posterior probability then only 0.21%
- Sample size required for 80% power at p = 0.05: n = 1.2M

VEGAS results candidate genes

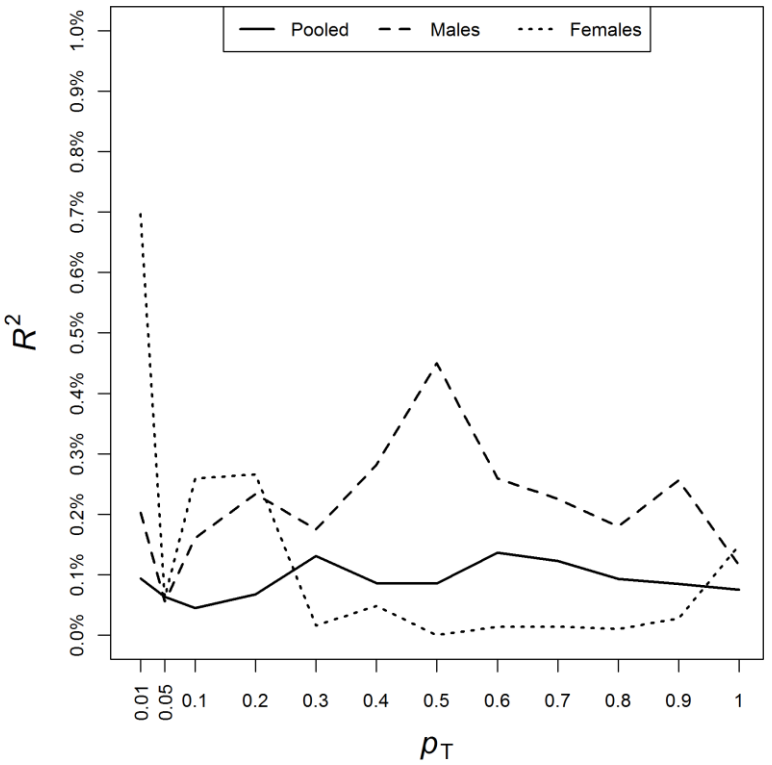
	***p*-value**	
Candidate gene	**Once**	**Serial**
ADORA2A	0.223	0.730
ADRA2A	0.003	0.871
COMT	0.128	0.095
DDC	0.411	0.345
DRD1	0.804	0.908
DRD2	0.276	0.034
DRD3	0.050	0.171
DRD4	0.238	0.668
DRD5	0.573	0.606
DYX1C1	0.643	0.663
HTR1B	0.371	0.812
HTR1E	0.904	0.273
HTR2A	0.518	0.745
KIAA0319 (*DYX2*)	0.142	0.980
ROBO1	0.940	0.885
SLC6A3 (*DAT1*)	0.398	0.792
SNAP25	0.254	0.648

- VEGAS results are based on the discovery stage meta-analyses
- 17 genes → Bonferroni-adjusted $p = 0.05 / 17 = 0.003$
- Different analyses for males, females, and two phenotypes → Bonferroni-adjusted $p = 0.05 / 102 = 4.9 \times 10^{-4}$

Prediction results



Once self-employment



Serial self-employment

- Predicting once/serial self-employment in STR using the discovery stage meta-analysis results
- Variance explained (Nagelkerke pseudo R^2 from logistic regression of phenotype on genetic score) vs. *p*-value threshold p_T for including SNPs in the genetic score calculation

Main lessons

- GREML results:
  - GWAS and prediction are in principle possible
- GWAS results:
  - Effects of individual SNPs are very low
  - Even larger sample sizes are needed for GWAS
- Challenges we faced:
  - Phenotype measurement
  - Entrepreneurship is a multifaceted, biologically distal outcome
  - Are results poolable across countries?
- Ways forward:
  - Larger N?
  - Endophenotypes?
  - New statistical approaches?