

Fifty Years of Twin Studies: A Meta-Analysis of the Heritability of Human Traits

Tinca JC Polderman^{1†}, Beben Benyamin^{2†}, Christiaan A. de Leeuw^{1,3}, Patrick F Sullivan^{4,5,6}, Arjen van Bochoven⁷, Peter M Visscher^{2,8‡}, Danielle Posthuma^{1,9*†}

¹ Department of Complex Trait Genetics, VU University, Center for Neurogenomics and Cognitive Research, Amsterdam, The Netherlands

² The University of Queensland, Queensland Brain Institute, Brisbane, Queensland 4072, Australia

³ Radboud University Nijmegen, Institute for Computing and Information Sciences, Nijmegen, The Netherlands

⁴ Center for Psychiatric Genomics, Department of Genetics, University of North Carolina, Chapel Hill, NC, USA

⁵ Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA

⁶ Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

⁷ Faculty of Sciences, VU University, Amsterdam, the Netherlands

⁸ The University of Queensland Diamantina Institute, The Translational Research Institute, Brisbane, Queensland, Australia

⁹ Department of Clinical Genetics, VU University Medical Center, Neuroscience Campus Amsterdam, The Netherlands

† These authors contributed equally to this work

‡ These authors contributed equally to this work

*Correspondence should be addressed to: Danielle Posthuma: Department of Complex Trait Genetics, VU University, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands. Phone: +31 20 598 2823, Fax: +31 20 5986926, d.posthuma@vu.nl

Word count: abstract: 150; main text: 3247

Display items: 5 (Tables 1-2, Figures 1-3)

Includes **Online Methods, Supplementary Materials** (incl. Tables S1-S27, Figs. S1-S12) and **Supplementary Tables** in excel S29-S33

1 **Abstract:** Despite a century of research in complex traits in humans, the relative importance and
2 specific nature of the influences of genes and environment on human traits remains controversial.
3 We report a meta-analysis of twin correlations and reported variance components for 17,804
4 traits from 2,748 publications including 14,558,903 partly dependent twin pairs, virtually all
5 published twin studies of complex traits. Estimates of heritability cluster strongly within
6 functional domains and across all traits the reported heritability is 49%. For a majority (69%) of
7 traits, the observed twin correlations are consistent with a simple and parsimonious model where
8 twin resemblance is solely due to additive genetic variation. The data are inconsistent with a
9 substantial influence of shared environment or non-additive genetic variation. This study
10 provides the most comprehensive analysis of the causes of individual differences in human traits
11 to date, and will guide future gene-mapping efforts. All results can be visualized at
12 <http://match.ctglab.nl>.

13 Insight into the nature of observed variation in human traits is important in medicine,
14 psychology, social sciences and evolutionary biology. It has gained new relevance with both the
15 ability to map genes for human traits and the availability of large, collaborative datasets to do so
16 on an extensive and comprehensive scale. Individual differences in human traits have been
17 studied for more than a century, yet the causes of human trait variation remain uncertain and
18 controversial. Specifically, the partitioning of observed variability into underlying genetic and
19 environmental sources and the relative importance of additive and non-additive genetic variation
20 are continually debated¹⁻⁵. Recent results from large-scale genome-wide association studies
21 (GWAS) show that many genetic variants contribute to complex trait variation and that effect
22 sizes are typically small^{6,7}. However, the sum of the variance explained by the detected variants
23 is much smaller than the reported heritability of the trait^{4,6-10}. This ‘missing heritability’ has led
24 some investigators to conclude that non-additive variation must be important^{4,11}. Although the
25 presence of gene-gene interaction has been demonstrated empirically^{5,12-17}, little is known about
26 its relative contribution to observed variation¹⁸.

27 In this study, our aim is two-fold. First, we analyze empirical estimates of the relative
28 contributions of genes and environment for virtually all human traits investigated in the past 50
29 years. Second, we assess empirical evidence for the presence and relative importance of non-
30 additive genetic influences on all human traits. We rely on classical twin studies, as the twin
31 design has been used widely to disentangle the relative contributions of genes and environment,
32 across a variety of human traits. The classical twin design is based on contrasting trait
33 resemblance of monozygotic (MZ) and dizygotic (DZ) twin pairs. MZ twins are genetically
34 identical and DZ twins are genetically full siblings. We show that for a majority (69%) of traits,
35 the observed statistics are consistent with a simple and parsimonious model where observed

36 variation is solely due to additive genetic variation. The data are inconsistent with a substantial
37 influence of shared environment or non-additive genetic variation. We also show that estimates
38 of heritability cluster strongly within functional domains and across all traits the reported
39 heritability is 49%. Our results are based on a meta-analysis of twin correlations and reported
40 variance components for 17,804 traits from 2,748 publications including 14,558,903 partly
41 dependent twin pairs, virtually all published twin studies of complex traits between 1958 and
42 2012. This study provides the most comprehensive analysis of the causes of individual
43 differences in human traits to date, and will guide future gene-mapping efforts. All results can be
44 visualized at <http://match.ctglab.nl>.

45

46 **RESULTS**

47 *The distribution of studied traits in twin studies is non-random across traits and countries*

48 We systematically retrieved published classical twin studies in which observed variation of
49 human traits was partitioned into genetic and environmental influences. For each study we
50 collected reported twin correlations for continuous traits and contingency tables for dichotomous
51 traits, estimates from genetic model fitting, and study characteristics (sample size, population,
52 age cohort and ascertainment scheme) (**Supplementary Table S1**). Investigated traits were
53 manually classified using the chapter and subchapter levels of the International Classification of
54 Functioning, Disability and Health (ICF) or the International Statistical Classification of
55 Diseases and Related Health Problems (ICD-10) (**Online Methods**). [The ICD10 and ICF](#)
56 [subchapter levels refer to actual diseases \(e.g. ‘Atopic Dermatitis’\) or traits \(e.g. ‘Temperament](#)
57 [and Personality Functions’\)](#). We identified 2,748 relevant twin studies, published between 1958
58 and 2012²⁰. Half of these were published after 2004, with sample sizes per study in 2012 around

59 1000 pairs (**Supplementary Table S2**). Each study could report on multiple traits measured in
60 one or several samples. These 2,748 studies reported on 17,804 traits. Twin subjects came from
61 39 different countries with a large proportion of studies (34%) based on US twin samples. The
62 continents of South-America (0.5%), Africa (0.2%) and Asia (5%) were heavily
63 underrepresented (**Fig. 1a,b, Supplementary Table S3**). The total number of studied twins was
64 14,558,903 partly dependent pairs, or 2,247,128 when correcting for reporting on multiple traits
65 per study. The majority of studies (59%) were based on the adult (aged 18-64) population,
66 although the sample sizes available for the elderly (aged 65+) population studies were the largest
67 (**Supplementary Table S4**). Authorship network analyses revealed that 61 communities of
68 authors wrote the 2,748 published studies. The 11 largest authorship communities contained >65
69 authors and could be mapped back to the main international twin registries, such as the Vietnam
70 Era Twin Registry, the Finnish Twin Cohort and the Swedish Twin Registry (**Supplementary**
71 **Figure S12**).

72 The investigated traits fell into 28 general trait domains. The distribution of traits evaluated in
73 twin studies was highly skewed with 62% focusing on traits classified under the *Psychiatric*,
74 *Metabolic* and *Cognitive* domains whereas traits classified under the *Developmental*, *Connective*
75 *Tissue*, and *Infection* domains together accounted for less than 1% of all investigated traits (**Fig.**
76 **1c, Supplementary Tables S5-S7**). The ten most investigated traits were *Temperament and*
77 *Personality Functions*, *Weight Maintenance Functions*, *General Metabolic Functions*,
78 *Depressive Episode*, *Higher-Level Cognitive Functions*, *Conduct Disorders*, *Mental and*
79 *Behavioural Disorders Due to Use of Alcohol*, *Anxiety Disorders*, *Height*, and *Mental and*
80 *Behavioural Disorders Due to Use of Tobacco*. Collectively these traits accounted for 59% of all
81 investigated traits.

82

83 *Across all traits, the reported relative contribution of genes and environment is equal*

84 We did not find evidence of systematic publication bias as a function of sample size (e.g. where
85 studies based on relatively small samples were only published when larger effects were reported)
86 (Fig. 1d, Supplementary Figure S1, Supplementary Tables S8-S11). We calculated weighted
87 averages of MZ and DZ twin correlations and of the reported estimates of the relative
88 contribution of genetic and environmental influences to investigated traits using a random effects
89 meta-analytic model to allow for heterogeneity across different studies. The meta-analyses of all
90 traits yielded an r_{MZ} of 0.64 and an r_{DZ} of 0.34. The reported heritability (h^2) across all traits was
91 0.49, and the reported estimate of shared environmental effects (c^2) was 0.17 (Table 1, Fig. 2a,b,
92 Supplementary Figure S6).

93

94 *Estimates of variance components cluster in functional domains*

95 Heritability estimates were found to cluster in functional domains, with the largest heritability
96 estimates for traits classified under the *Ophthalmological* domain, followed by the *Ear-Nose-*
97 *Throat, Dermatological* and *Skeletal* domains. The lowest heritability estimates were for traits in
98 the *Environment, Reproduction* and *Social Values* domains (Fig. 2d, Supplementary Table
99 S17). All weighted averages of h^2 across >500 distinct traits had a mean greater than zero
100 (Supplementary Tables S18-S20, S31, S32). The lowest reported heritability for a specific trait
101 was for *Gene Expression* with an estimate of $h^2=0.05$ (SE=0.03), and a c^2 of 0.74 (SE=0.03), but
102 note these estimates are based on only 20 entries. The largest influence of c^2 was found for traits
103 in the *Cell* domain (0.67, SE=0.05), followed by traits in the *Infection* (0.35, SE=0.05),

104 *Hematological* (0.32, SE=0.09), *Endocrine* (0.32, SE=0.05), *Reproduction* (0.32, SE=0.06),
105 *Social Values* (0.27, SE=0.03), *Environment* (0.25, SE=0.02), and *Skeletal* (0.25, SE=0.02)
106 domains (**Fig. 2d, Supplementary Table S17**).

107

108 *Heterogeneity of twin correlations across sex and age*

109 Across all traits, the weighted averages of twin correlations and reported h^2 and c^2 did not show
110 evidence of heterogeneity across sex, although there was some evidence for a lower correlation
111 in opposite sex twin pairs versus same sex dizygotic pairs (**Table 1, Supplementary**
112 **Information, section 5.1**). The data showed a decrease in MZ and DZ resemblance after
113 adolescence and an accompanying decrease in estimates of both h^2 and c^2 (**Fig. 2c,**
114 **Supplementary Table S16**).

115 In the top 20 most investigated traits for twin correlations the weighted estimates did not show
116 consistent evidence for heterogeneity across sex, with r_{MZM} and r_{MZF} as well as r_{DZM} and r_{DZF}
117 remarkably similar across the majority of the top 20 investigated specific traits (**Fig. 3**), although
118 for several traits the opposite sex twin correlations were lower than the same sex estimates,
119 mostly after age 11 (e.g. for *Weight Maintenance Functions*, *Functions of Brain and Mental,*
120 *Behavioral Disorders due to the Use of Alcohol*, and *Mental and Behavioral Disorders due to the*
121 *Use of Tobacco*). Heterogeneity of weighted twin correlations across age was more prominent
122 than heterogeneity across sex (**Fig. 3**). For example when considering the r_{MZ} we note that for
123 most of the top 20 investigated traits the estimate tends to decrease with age, especially after
124 adolescence, which was generally mirrored in the r_{DZ} (**Fig. 3**).

125

126 *Per study, variance components model fitting and selection leads to an underestimation of the*
127 *h^2*

128 Falconer's equations can be used to calculate \hat{h}^2 and \hat{c}^2 based on twin correlations¹⁸. The
129 equation for \hat{h}^2 is $2 \times (r_{MZ} - r_{DZ})$, and for \hat{c}^2 it is $2 \times r_{DZ} - r_{MZ}$. When these are applied to the weighted
130 averages of r_{MZ} and r_{DZ} we find an \hat{h}^2 estimate of $2 \times (0.64 - 0.34) = 0.60$ and a \hat{c}^2 estimate of
131 $2 \times 0.34 - 0.64 = 0.04$ (**Table 1, Supplementary Figure S7**). We note that the estimated \hat{h}^2 based on
132 twin correlations is larger than the weighted average of reported h^2 . As a consequence, the
133 estimated \hat{c}^2 based on twin correlations is lower than the weighted average of the reported c^2
134 component. To test whether this discrepancy was due to a bias in studies reporting only twin
135 correlations or only variance components, we conducted the meta-analysis only on studies
136 reporting both. This yielded similar estimates with a similar discrepancy (**Supplementary Table**
137 **S13**), ruling out the explanation that twin correlations may have been reported on traits that
138 happened to be more heritable than traits for which the estimates of variance components were
139 reported. Through theory we show that such a discrepancy can arise when the individual studies
140 represent a mix of traits that follow a pattern of $r_{MZ} > 2r_{DZ}$ and $r_{MZ} < 2r_{DZ}$, and where the choice
141 of fitting a model that includes shared environment or non-additive genetic influences is based
142 on the observed pattern of twin correlations (see **Supplementary Information, section 5.2**).
143 More specifically, because c^2 and non-additive genetic influences cannot be estimated
144 simultaneously from twin correlations an 'ACE' model (Additive genetic, Common
145 environmental and Error or non-shared environmental influences) is fitted to the data if $2r_{DZ} -$
146 $r_{MZ} > 0$. On the other hand, if $2r_{DZ} - r_{MZ} < 0$, an 'ADE' model, including non-additive genetic
147 instead of common environmental influences, is selected. This leads to sampling bias in the
148 estimate of h^2 from the full model. We show (**Supplementary Table S14**) that such per-study

149 choices cause bias and can lead to a 10% downward bias of reported estimates of h^2 compared to
150 those based on twin correlations, consistent with the observed discrepancy between our meta-
151 analysis of variance component estimates calculated from twin correlations and the reported
152 variance components.

153

154 *The overall pattern of twin correlations is consistent with a model where all genetic variance is*
155 *additive and all environmental variance is non-shared*

156 There may be many causes of similarities and differences within MZ and DZ pairs and typically
157 these are interpreted in terms of (non-) additive genetic and (non-) shared environmental
158 influences¹⁹. Yet, there are essentially only two estimable and testable variance components of
159 interest in the twin design. Therefore, inference from classical twin studies on all underlying
160 unobserved sources of variation that lead to the resemblance between relatives is limited.
161 However, there are two simple and parsimonious hypotheses that can be tested across traits from
162 estimated correlation coefficients of MZ pairs (r_{MZ}) and DZ pairs (r_{DZ}). The first is that the MZ
163 and DZ population correlations (ρ_{MZ} and ρ_{DZ}) are the same, consistent with twin resemblance
164 being solely due to non-genetic factors. The second is a two-fold ratio of ρ_{MZ} and ρ_{DZ} , consistent
165 with twin resemblance being solely due to additive genetic variation. An across-trait consistency
166 with either of these hypotheses is not a proof of these simple models but would provide an
167 extremely parsimonious model against which other experimental designs (e.g. DNA-sequence
168 based) should be tested. For the vast majority (84%) of traits, we find that MZ correlations are
169 larger than DZ correlations. Using the weighted estimates of r_{MZ} and r_{DZ} across all traits, we
170 show that the average $2r_{DZ}-r_{MZ} = .04$ (SE=.01) (**Table 1**), which is very close to a two-fold
171 difference in the MZ to DZ correlation. The proportion of single studies in which the pattern of

172 twin correlations was consistent with the null hypothesis that $2r_{DZ}=r_{MZ}$ was 69%. This observed
173 pattern of twin correlations is consistent with a simple and parsimonious underlying model of the
174 absence of environmental effects shared by twin pairs and the presence of genetic effects that are
175 entirely due to additive genetic variation (**Table 2**). This remarkable fitting of the data with a
176 simple mode of family resemblance is inconsistent with the hypothesis that a substantial part of
177 human trait variation is due to shared environmental variation or to substantial non-additive
178 genetic variation.

179

180 *For most specific traits the empirical data is consistent with a model where all genetic*
181 *variance is additive*

182 Although across all traits 69% of studies show a pattern of MZ and DZ twin correlations
183 consistent with an r_{MZ} that is exactly twice the r_{DZ} , this finding is not necessarily representative
184 of the majority of studies in functional domains or for every specific trait (i.e., at ICD10/ICF
185 subchapter level). We thus calculated the proportion of studies consistent with $2r_{DZ}=r_{MZ}$ within
186 functional domains and for each specific trait, and found that traits consistent with this
187 hypothesis tend to cluster in specific functional domains. A pattern of twin correlations
188 consistent with $2r_{DZ}=r_{MZ}$ was most prominent for traits included in the *Neurological*, *Ear-Nose-*
189 *Throat*, *Cardiovascular* and *Ophthalmological* domains, with 99.5%, 97%, 95% and 87% of
190 studies being consistent with a model where all resemblance was entirely due to additive genetic
191 variance. In only three of 28 general trait domains most studies were inconsistent with this
192 model. These domains were *Activities* (35%), *Reproduction* (44%), and *Dermatological* (45%)
193 (**Table 2** and **Supplementary Table S24**). Of 59 specific traits (i.e. ICD10/ICF subchapter
194 classification) for which we had sufficient information to calculate the proportion of studies

195 consistent with $2r_{DZ}=r_{MZ}$, 21 traits showed a proportion less than 0.50, whereas for the remaining
196 38 traits the majority of individual studies were consistent with $2r_{DZ}=r_{MZ}$ (**Supplementary Table**
197 **S26**). Of the top 20 most investigated specific traits, we found that for 12 traits the majority of
198 individual studies were consistent with a model where variance was solely due to additive
199 genetic variance and non-shared environmental variance, while the pattern of MZ and DZ twin
200 correlations was inconsistent with this model for eight traits, suggesting that apart from additive
201 genetic influences and non-shared environmental influences, either or both non-additive genetic
202 influences and shared environmental influences are needed to explain the observed pattern of
203 twin correlations (**Table 2 and Supplementary Table S26**). These eight traits were *Conduct*
204 *disorders*, *Height*, *Higher Level Cognitive Functions*, *Hyperkinetic disorder*, *Mental and*
205 *Behavioral Disorders due to the use of Alcohol*, *Mental and Behavioral Disorders due to the use*
206 *of Tobacco*, *Other Anxiety Disorders*, and *Weight Maintenance Functions*. For all eight traits,
207 meta-analyses on reported variance components resulted in a weighted estimate of reported
208 shared environmental influences that was statistically different from zero (**Supplementary**
209 **Table S32**). Comparison of weighted twin correlations for these specific traits resulted in
210 positive estimates of $2r_{DZ}-r_{MZ}$, except for *Hyperkinetic Disorders* where $2r_{DZ}-r_{MZ}$ was -0.13
211 (SE=0.03), based on 144 individual reports and 207,589 twin pairs), which suggests the
212 influence of non-additive genetic variation for this trait, or any other source of variation that
213 leads to a disproportionate similarity among monozygotic twin pairs.

214

215 **DISCUSSION**

216 We have conducted a meta-analysis of virtually all twin studies published in the past 50 years, on
217 a wide range of traits and reporting on more than 14 million twin pairs across 39 different

218 countries. Our results provide compelling evidence that all human traits are heritable: not one
219 trait had a weighted heritability estimate of zero. The relative influence of genes and
220 environment is not randomly distributed across all traits but clusters in functional domains. In
221 general, we showed that reported estimates of variance components from model fitting can
222 underestimate the true trait heritability, when compared with the heritability based on the twin
223 correlations. Roughly two-thirds of traits show a pattern of MZ and DZ twin correlations that is
224 consistent with a simple model whereby trait resemblance is solely due to additive genetic
225 variation. This implies that for the majority of complex traits causal genetic variants can be
226 detected using a simple additive genetic model.

227 Approximately one-third of traits did not follow the simple pattern of a two-fold ratio in MZ and
228 DZ correlations. For these traits, a simple additive genetic model does not sufficiently describe
229 the population variance. An incorrect assumption about narrow sense heritability (the proportion
230 of total phenotypic variation due to additive genetic variation) can lead to a mismatch between
231 results from gene finding studies and prior expectations. If the pattern of twin correlations is
232 consistent with a substantial contribution from shared environmental factors, like we find for
233 *Conduct Disorders, Religion and Spirituality, and Education*, then gene-mapping studies may
234 yield disappointing results. If the cause of a departure from a simple additive genetic model is
235 the existence of non-additive genetic variation, as is for example suggested by the average twin
236 correlations for *Recurrent Depressive Disorder, Hyperkinetic Disorders, and Atopic Dermatitis*,
237 then it may be tempting to fit non-additive models in gene mapping studies (e.g. GWAS or
238 sequencing studies). However, the statistical power of such scans is extremely low due to the
239 many non-additive models that can be fitted (e.g. within-locus dominance versus between locus
240 additive by additive effects) and the penalty incurred by multiple testing. Our current results

241 signal traits for which an additive model cannot be assumed. In most of these traits DZ twin
242 correlations are higher than half the MZ correlations, suggesting that shared environmental
243 effects are causing the deviation from a simple additive genetic model. Yet, data from twin pairs
244 only does not provide sufficient information to resolve the actual causes of the deviation from a
245 simple additive genetic model. More detailed studies may reveal the likely causes of such
246 deviation, and may as such uncover epidemiological or biological factors that drive family
247 resemblance. To make stronger inferences about the causes underlying the resemblance between
248 relatives for traits that deviate from the additive genetic model, additional data are required, for
249 example from large population samples with extensive phenotypic and DNA sequence
250 information, detailed measures of environmental exposures, and larger pedigrees including non-
251 twin relationships.

252 We note that our inference is based on twin studies published between 1950 and 2012 and
253 generally applies to complex traits and not necessarily generalizes to Mendelian subtypes of
254 traits. Most Mendelian traits are rare in the population and are therefore not studied by twin
255 researchers because they cannot ascertain enough affected twin pairs to estimate genetic
256 parameters reliably. In the rare case that they were available, the Mendelian subtypes were
257 analyzed together with the subtypes of the same trait that are due to common causes. In that case
258 our inference would be biased away from our main result because Mendelian diseases tend to be
259 dominant or recessive, not additive. In addition, there may be heterogeneity in measurement
260 errors between studies for the same trait and between traits. A test-retest correlation would
261 quantify measurement error when contrasted with a correlation between MZ twins but few twin
262 studies report such correlations in the same papers that estimate heritability.

263 Our results provide the most comprehensive empirical overview of the relative contributions of
264 genes and environment to all human traits that have been studied in twins to date, which can
265 guide and serve as a reference for future gene-mapping efforts.

266

267 **References and Notes**

268

- 269 1. Moore, J.H. Analysis of gene-gene interactions. *Curr. Protoc. Hum. Genet.* Chapter 1, Unit
270 1.14 (2004).
- 271 2. Hill, W.G., Goddard, M.E. & Visscher, P.M. Data and theory point to mainly additive
272 genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).
- 273 3. Traynor, B.J. & Singleton, A.B. Nature versus nurture: death of a dogma, and the road ahead.
274 *Neuron* **68**, 196–200 (2010).
- 275 4. Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S. The mystery of missing heritability:
276 Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1193–
277 1198 (2012).
- 278 5. Phillips, P.C. Epistasis--the essential role of gene interactions in the structure and evolution
279 of genetic systems. *Nat. Rev. Genet.* **9**, 855–867 (2008).
- 280 6. Visscher, P.M., Brown, M.A., McCarthy, M.I. & Yang, J. Five years of GWAS discovery.
281 *Am. J. Hum. Genet.* **90**, 7–24 (2012).
- 282 7. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–
283 753 (2009).
- 284 8. Stranger, B.E., Stahl, E.A. & Raj, T. Progress and promise of genome-wide association
285 studies for human complex trait genetics. *Genetics* **187**, 367–383 (2011).
- 286 9. Maher, B. Personal genomes: The case of the missing heritability. *Nature* **456**, 18–21 (2008).
- 287 10. Eichler, E.E. Flint, J., Gibson, G., Kong, A., Leal, S. M., *et al.* Missing heritability and
288 strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* **11**, 446–
289 450 (2010).

- 290 11. Nelson, R.M., Pettersson, M.E. & Carlborg, Ö. A century after Fisher: time for a new
291 paradigm in quantitative genetics. *Trends Genet. TIG* **29**, 669–676 (2013).
- 292 12. Barker, J.S. Inter-locus interactions: a review of experimental evidence. *Theor. Popul. Biol.*
293 **16**, 323–346 (1979).
- 294 13. Cockerham, C.C. An extension of the concept of partitioning hereditary variance for analysis
295 of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882 (1954).
- 296 14. Cockerham, C.C. in *Statistical Genetics and Plant Breeding* 53–94 (Nat. Acad. Sci. Nat. Res.
297 Council Publ., 1963).
- 298 15. Kempthorne, O. On the covariances between relatives under selfing with general epistacy.
299 *Proc. R. Soc. Lond. Ser. B Contain. Pap. Biol. Character R. Soc. G. B.* **144**, 100–108 (1956).
- 300 16. Crow, J.F. & Kimura, M. *An Introduction To Population Genetics Theory*. (Harper and Row,
301 1970).
- 302 17. Carlborg, O. & Haley, C.S. Epistasis: too often neglected in complex trait studies? *Nat. Rev.*
303 *Genet.* **5**, 618–625 (2004).
- 304 18. Falconer, D.S. & Mackay, T.F.C. *Quantitative Genetics*. (Longman Group Ltd, 1996).
- 305 19. Lynch, M. & Walsch, B. *Genetics and Analysis of Quantitative Traits*. (Sinauer Associates,
306 1998).
- 307 20. A full reference list of al 2748 studies included in the meta-analyses is provided in

308 **Supplementary Table S33**

309

310 **Supplementary Information** is available in the online version of the paper.

311

312 **Acknowledgements:** This work was funded by The Netherlands Organization for Scientific
313 Research (NWO VICI 453-14-005, 645-000-003), the Australian Research Council

314 (DP130102666) and by the Australian National Health and Medical Research Council
315 (APP613601). We would like to thank Marijke Frantsen, Mark Patrick Roeling, Rebecca Lee,
316 and Daniela Marie DeCristo for their contribution to collecting full texts of selected twin studies
317 and data entry.

318

319 **Author Contributions:** B.B., D.P., P.F.S, and P.V. performed the analyses. D.P. conceived the
320 study. D.P., T.P., and P.V. designed the study. D.P. and T.P. collected and entered the data. D.P.
321 and P.F.S. categorized traits according to standard classifications; AvB. and CdL. checked data
322 entries and checked and wrote statistical scripts. AvB. programmed the web-tool. D.P. T.P. and
323 P.V. wrote the paper. All authors discussed the results and commented on the paper. T.P. and
324 B.B, and P.V. and D.P. contributed equally to this work.

325

326 **Author Information:** The data reported in this manuscript can be viewed in the web tool
327 MaTCH: Meta-Analyses of Twin Correlations and Heritability: <http://match.ctglab.nl>. Reprints
328 and permissions information is available at www.nature.com/reprints. The authors declare no
329 competing financial interests. Correspondence and requests for materials should be addressed to
330 d.posthuma@vu.nl.

331

332 **Table 1. Weighted means of twin correlations and variance components across all human**
 333 **traits investigated in a classical twin study and published between 1958-2012**

	Est.	SE	NTraits	Npairs
r_{MZ}	0.64	0.0024	9568	2,563,627
r_{MZM}	0.62	0.0037	4518	1,070,962
r_{MZF}	0.63	0.0035	4360	1,171,841
r_{DZ}	0.34	0.0031	5220	2,606,252
r_{DZSS}	0.35	0.0030	6108	1,752,952
r_{DZM}	0.32	0.0034	4412	1,039,238
r_{DZF}	0.34	0.0037	4255	1,068,562
r_{DOS}	0.30	0.0047	2342	898,610
h^2	0.49	0.0042	2929	4,341,721
h^2_{SS}	0.47	0.0047	1795	1,187,837
h^2_M	0.47	0.0054	2095	1,732,622
h^2_F	0.47	0.0050	1957	1,539,582
c^2	0.17	0.0040	2771	4,272,318
c^2_{SS}	0.19	0.0055	1769	1,185,116
c^2_M	0.16	0.0042	1988	1,519,148
c^2_F	0.17	0.0047	1925	1,516,192
$2(r_{MZ}-r_{DZ})$	0.59	0.0078	9568	5,169,879
$2(r_{MZ}-r_{DZ})_{SS}$	0.58	0.0077	9568	4,316,578
$2(r_{MZ}-r_{DZ})_M$	0.59	0.0102	4518	2,110,200
$2(r_{MZ}-r_{DZ})_F$	0.57	0.0102	4360	2,240,403
$2r_{DZ}-r_{MZ}$	0.04	0.0066	9568	5,169,879
$2r_{DZ}-r_{MZ}_{SS}$	0.05	0.0065	9568	4,316,578
$2r_{DZ}-r_{MZ}_M$	0.02	0.0078	4518	2,110,200
$2r_{DZ}-r_{MZ}_F$	0.06	0.0082	4360	2,240,403

334 r =correlation; MZ=monozygotic twins; DZ=dizygotic twins; M=males; F=females; SS=same sex
 335 pairs only; DOS=DZ opposite sex; h^2 =heritability; c^2 =proportion of variance due to share
 336 environmental variation; Est.=estimate based on random effects meta-analysis; SE=standard
 337 error; NTraits=number of investigated traits; Npairs=number of dependent twin pairs. The pairs
 338 are not independent as the same or an overlapping sample of twins may have been used for
 339 multiple traits and across multiple studies.

340 **Table 2. Proportion (π_0) of studies that are consistent with a model where trait resemblance is**
 341 **solely due to additive genetic variation, and weighted means of twin correlations, for the main**
 342 **trait domains and the top 20 investigated traits**

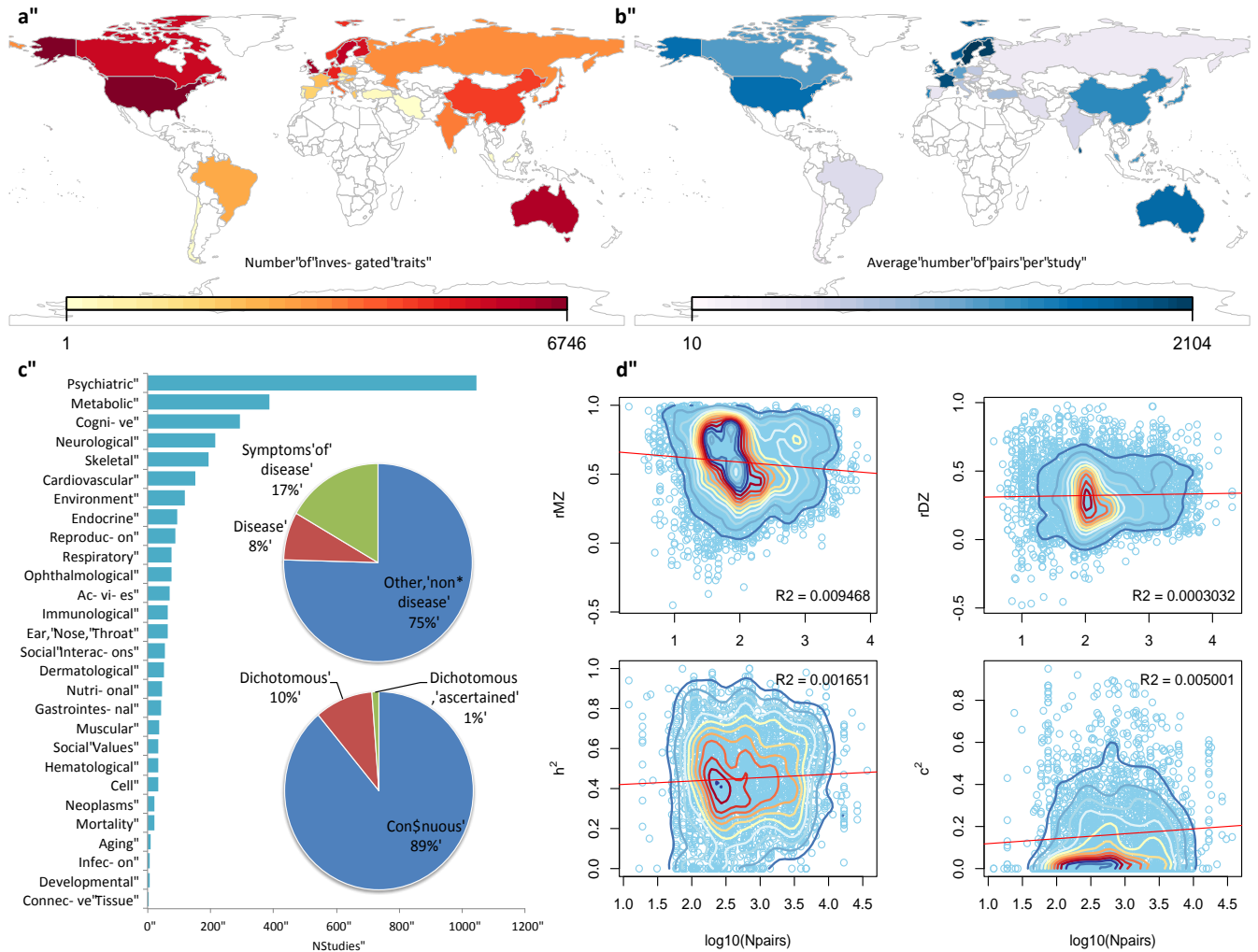
	π_0				r_{MZ}		r_{DZ}			
	NTraits	Est.	Est.	SE	NTraits	Npairs	Est.	SE	NTraits	Npairs
All traits	5185	0.69	0.64	0.0024	9568	2,563,628	0.34	0.0031	5220	2,606,252
General trait domains										
Activities	62	0.35	0.57	0.0185	118	58,227	0.34	0.0217	63	55,864
Cardiovascular	267	0.95	0.56	0.0080	380	41,669	0.29	0.0096	268	25,544
Cell	54	0.59	0.72	0.0219	72	3,188	0.52	0.0428	54	1,667
Cognitive	450	0.57	0.65	0.0072	931	288,867	0.37	0.0103	454	304,720
Dermatological	74	0.45	0.73	0.0251	109	19,509	0.40	0.0175	75	23,245
Ear, Nose, Throat	165	0.97	0.76	0.0134	200	27,882	0.33	0.0155	172	14,222
Endocrine	108	0.69	0.56	0.0171	162	10,112	0.39	0.0218	110	9,140
Environment	145	0.50	0.55	0.0136	295	120,606	0.40	0.0167	145	99,137
Gastrointestinal	32	0.59	0.55	0.0243	64	10,982	0.27	0.0281	39	28,431
Hematological	19	0.65	0.76	0.0229	50	5,541	0.56	0.0316	19	3,218
Immunological	230	0.67	0.61	0.0116	280	18,051	0.36	0.0128	231	36,075
Metabolic	464	0.60	0.75	0.0049	912	210,189	0.41	0.0078	464	197,921
Neurological	702	1.00	0.68	0.0048	1751	129,076	0.29	0.0058	705	89,103
Nutritional	110	0.72	0.48	0.0157	205	75,751	0.22	0.0146	110	79,188
Ophthalmological	106	0.87	0.73	0.0165	199	26,139	0.39	0.0171	106	16,189
Psychiatric	1778	0.62	0.55	0.0044	2865	1,232,382	0.31	0.0050	1781	1,374,817
Reproduction	16	0.44	0.77	0.0336	34	12,130	0.33	0.0633	16	27,879
Respiratory	125	0.74	0.70	0.0183	184	34,443	0.33	0.0190	127	51,150
Skeletal	190	0.51	0.83	0.0077	395	111,282	0.50	0.0121	191	113,080
Social Interactions	24	0.63	0.34	0.0168	146	43,501	0.27	0.0411	24	22,764
Social Values	45	0.69	0.49	0.0297	120	52,492	0.41	0.0619	45	28,071
Top 20 investigated traits for r_{MZ} and r_{DZ}										
Blood Pressure Funct.	110	0.93	0.58	0.0095	179	20,621	0.31	0.0125	110	11,620
Conduct Dis.	216	0.41	0.66	0.0091	289	147,974	0.41	0.0096	216	192,651
Depressive Episode	115	0.60	0.45	0.0136	173	98,315	0.25	0.0153	115	121,936
Endocrine Gland Funct.	92	0.72	0.54	0.0170	139	8,533	0.38	0.0251	92	7,295
Food	110	0.72	0.48	0.0157	205	75,751	0.22	0.0146	110	79,188
Funct. of Brain	594	0.99	0.68	0.0062	1010	69,722	0.29	0.0064	594	58,951
General Metab. Funct.	219	0.69	0.68	0.0074	462	62,108	0.37	0.0101	219	58,338
Heart Funct.	140	1.00	0.53	0.0095	174	15,070	0.27	0.0111	140	11,109
Height	87	0.29	0.91	0.0045	128	53,076	0.54	0.0079	87	68,358
High-L. Cognitive Funct.	188	0.44	0.71	0.0087	419	152,197	0.44	0.0156	188	158,626
Hyperkinetic Dis.	100	0.37	0.65	0.0130	144	86,450	0.26	0.0159	100	121,139
Imm. System Funct.	223	0.67	0.61	0.0118	276	16,703	0.36	0.0131	223	32,964
Ment. Beh. Dis. Alc.	100	0.36	0.63	0.0152	158	94,477	0.41	0.0199	101	94,196
Ment. Beh. Dis. Tob.	70	0.47	0.72	0.0164	110	51,102	0.47	0.0220	72	34,186
Other Anxiety Dis.	145	0.29	0.55	0.0131	191	105,902	0.33	0.0157	145	153,730
Spec. Personal. Dis.	140	0.93	0.45	0.0092	162	41,460	0.23	0.0073	140	33,681
Structure of the Eyeball	86	0.91	0.73	0.0221	121	19,276	0.37	0.0190	86	13,580
Structure of Mouth	117	0.89	0.82	0.0103	127	7,769	0.40	0.0118	119	8,493
Temp. Pers. Funct.	568	0.84	0.47	0.0079	1134	334,190	0.23	0.0105	568	296,114
Weight Maint. Funct.	215	0.48	0.81	0.0051	391	141,152	0.44	0.0102	215	134,867

343 General trait domain categories with < 10 entries for π_0 were excluded. Alc. = Alcohol; Dis.=
 344 Disorders; High-L. = Higher Level; Imm. = Immunological; Funct. = Functions; Maint.=
 345 Maintenance; Metab.= Metabolic; Ment. Beh. = Mental and Behavioural; Spec. Personal.= Specific

346 Personality; Temp. Pers. =Temperament and Personality; Tob. = Tobacco; other abbreviations, see
347 Table 1. The top 20 investigated traits are conditional on the reporting of r_{MZ} and r_{DZ} .

348 **Figures and legends**

349 **Fig. 1. Distribution of investigated traits in virtually all twin studies published between**
 350 **1958 and 2012**

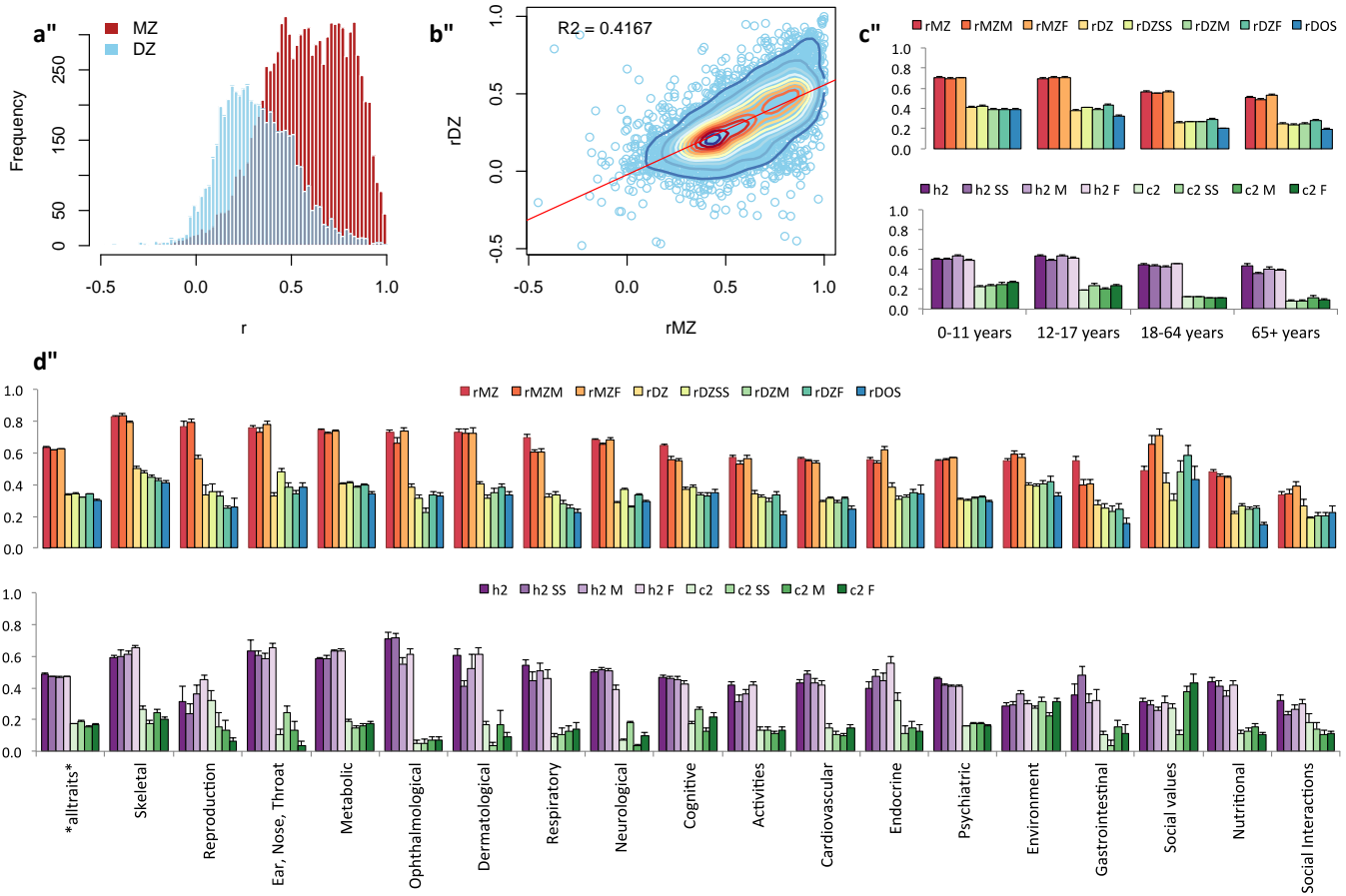


351 **(a)** The number of investigated traits in classical twins studies across all countries. **(b)** The
 352 average number of included pairs per study across countries. **(c)** The number of investigated
 353 traits according to functional trait domain and trait characteristic (inset). **(d)** MZ and DZ twin
 354 correlations and reported estimates of h^2 and c^2 as a function of sample size. Contour lines
 355 indicate the density of the data in that region. The lines are ‘heat’ colored from blue to red,
 356 indicating increasing data density.
 357

358

359

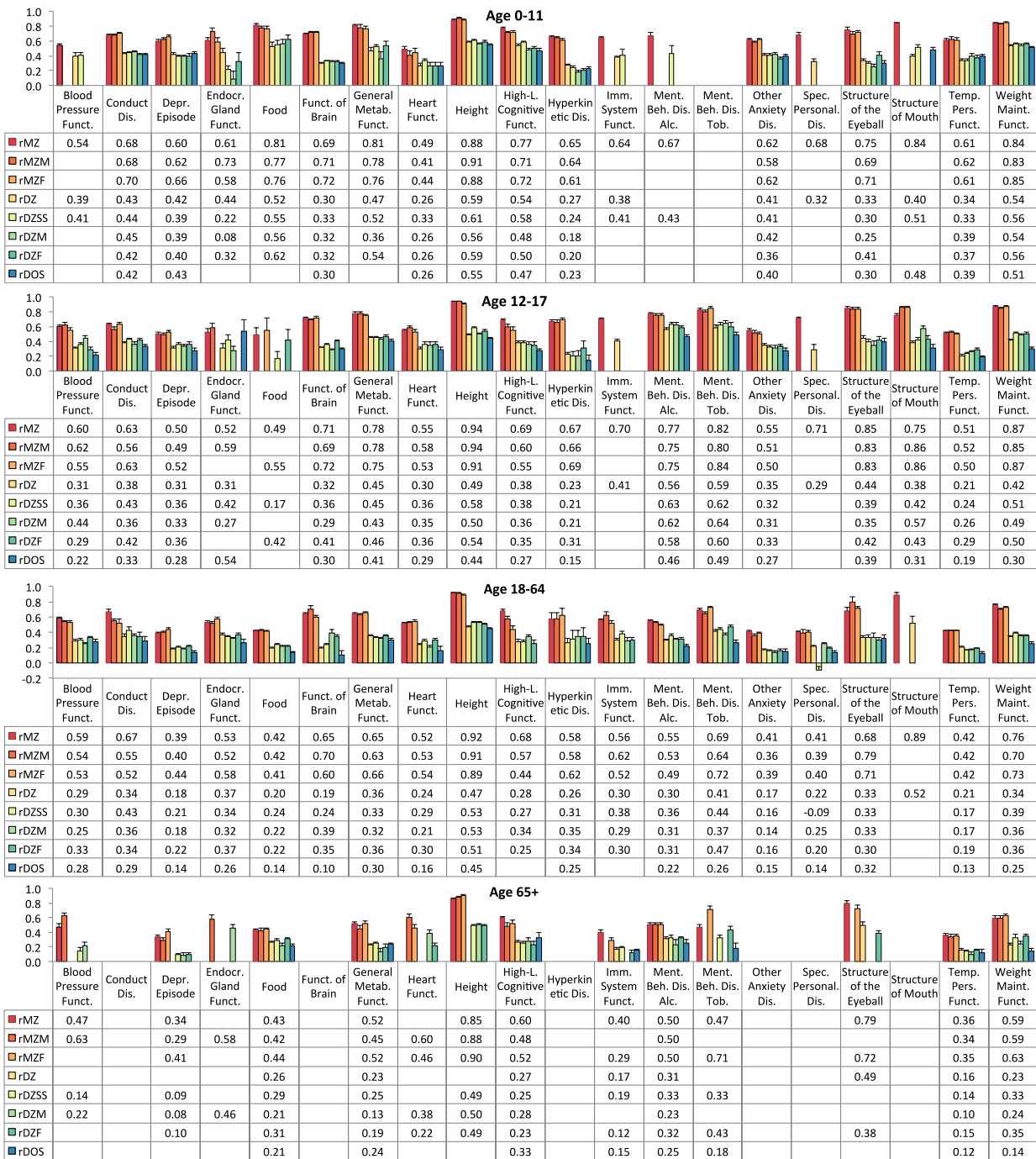
360 **Fig 2. Twin correlations and heritabilities for all human traits**



361

362 (a) Distribution of r_{MZ} and r_{DZ} across traits investigated in 2,748 twin studies published between
 363 1958 and 2012. The r_{MZ} is based on 9,568 traits and 2,563,628 partly dependent pairs; r_{DZ} is
 364 based on 5,220 traits and 2,606,252 partly dependent pairs (see **Table 1**). (b) Relation between
 365 the r_{MZ} and r_{DZ} , using all 5,185 traits for which both were reported. (c) Random effect meta-
 366 analytic estimates of twin correlations (top) and reported variance components (bottom) across
 367 all traits separate for four age cohorts. Error bars denote standard errors. (d) Random effect
 368 meta-analytic estimates of twin correlations (top) and reported variance components (bottom)
 369 across all traits and within functional domains for which data on all correlations and variance
 370 components was available. Error bars denote standard errors. For data Tables see Supplementary
 371 Information.
 372

373 **Fig. 3 Twin correlations for the top 20 most investigated specific traits across age and sex**



374
 375 Alc. = Alcohol; Dis.= Disorders; High-L. = Higher Level; Imm. = Immunological; Funct. =
 376 Functions; Maint.= Maintenance; Metab.= Metabolic; Ment. Beh. = Mental and Behavioural;
 377 Spec. Personal.= Specific Personality; Temp. Pers. =Temperament and Personality; Tob. =

378 Tobacco; other abbreviations, see Table 1. The top 20 investigated traits are conditional on the
379 reporting of r_{MZ} and r_{DZ} . Empty cells denote insufficient information available to calculate
380 weighted estimates; error bars denoted standard errors. We note that estimates and Figures for *all*
381 specific traits are available in the online webtool MaTCH.

382 **Online Methods**

383 **Identifying relevant studies**

384 We searched PubMed for all studies published between Jan 1st, 1900 and December 31st, 2012
385 that provided twin correlations, concordance rates, a heritability (h^2), or environmentability (c^2)
386 estimate, based on monozygotic (MZ) and dizygotic (DZ) twin data. The following search term
387 was used in PubMed:

388 *("english"[Language] AND ("1900/01/01"[Date - Publication] : "2012/12/31"[Date - Publication]) AND twin AND
389 "journal article"[Publication Type] AND "humans"[Filter] AND (heritability[Title/Abstract] OR "genetic
390 influence"[Title/Abstract] OR "environmental influence"[Title/Abstract] OR "genetic factors"[Title/Abstract] OR
391 "environmental factors"[Title/Abstract]) AND "journal article"[Publication Type]) NOT review[Title] NOT
392 review[Publication Type].*

393 The search was run on 31/01/2013 and again on 29/04/2013, which yielded an additional 44
394 publications, with the difference likely due to keywords or tags that had been added to
395 publications in PubMed in the intermittent period.

396 The last PubMed search yielded 4,388 unique studies. From these we deleted studies that were
397 not relevant for the current purpose using the following exclusion criteria:

- 398 - Studies in which only MZ twins were included (including discordant twin case studies)
- 399 - Studies in which no heritability estimates, twin correlations or concordances were
400 provided
- 401 - Review studies
- 402 - Meta-analysis of other studies
- 403 - Multivariate studies that provided information on completely overlapping traits and
404 samples with previously published univariate studies

405 - Some studies investigating h^2 of the brain (e.g. voxel-based brain measures) were not
406 included for practical purposes. These studies typically present their results in graphs
407 with color-coded point estimates of heritability mapped onto the brain. Such estimates
408 could not be quantified and thus were not entered into the database.

409 From the remaining 2,748 studies we were able to retrieve full texts from all but five (99.8%). Of
410 the studies without full-text availability, we included available relevant information based on the
411 abstract.

412

413 **Primary information obtained from each study**

414 From every study we retrieved basic information on PubMed ID, authors, the trait as named in
415 the study, and the year of publication. All information was entered into a database. In addition
416 the following information was retrieved:

417

418 - *The country of origin of the study population:* We used standard ISO country names and
419 where possible data entry was done separately for each country investigated in the study.
420 Note that in studies where pooled estimates across different countries were provided, yet
421 sample sizes were available for each country separately, we entered information
422 separately for each country.

423 - *The age group of the study population:* The study population was classified into four age
424 cohorts based on the average age of the population included in the study, as: age > 0 and
425 < 12 ; age ≥ 12 and < 18 ; age ≥ 18 and < 65 ; and age ≥ 65 .

426 - *The MZ and DZ twin correlations:* Twin correlations were entered as provided in the
427 study, and could be calculated as intra-class, Pearson, polychoric or tetrachoric

428 correlations, or based on least squares or maximum likelihood estimates. When available
429 we entered the twin correlations separated for males and females (i.e., MZ males
430 (MZM), MZ females (MZF), DZ males (DZM), DZ females (DZF), DZ opposite sex
431 (DOS)). If correlations were not available for males and females separately, we entered
432 the MZall and DZall correlations, i.e. based on both sexes. In cases where it was clear
433 that the DZ correlation was based on same-sex twins only, we entered the DZ same-sex
434 (DZSS) correlation.

435 - *The estimates of the heritability (h^2) and shared environmental component (c^2), under*
436 *the full ACE (or ADE) model:* We entered ‘ h^2_FULL ’ and ‘ c^2_FULL ’, based on
437 estimates under the full ACE (including additive genetic, shared and non-shared
438 environmental influences) or ADE (including additive and non-additive genetic and non-
439 shared environmental influences) model. When an ACE model was fitted, the estimate
440 for A was entered in ‘ h^2_FULL ’ and the estimated for C in ‘ c^2_FULL ’. When an ADE
441 model was fitted, the estimates of A and D were summed and entered for ‘ h^2_FULL ’
442 and zero was entered for ‘ c^2_FULL ’. When estimates were provided separately for
443 males and females, they were entered separately; otherwise they were entered for the
444 total sample of males and females. In the case of multivariate analyses, univariate
445 estimates were always preferred to allow comparison across studies.

446 - *The estimates of the heritability (h^2) and shared environmental component (c^2), under*
447 *the best fitting ACE (or ADE) model:* We entered ‘ h^2_BEST ’ and ‘ c^2_BEST ’, based on
448 estimates under the best fitting ACE or ADE model as provided in the study. When an
449 ACE model was the best fitting model, the estimate for A was entered in ‘ h^2_BEST ’ and
450 the estimate for C in ‘ c^2_BEST ’. When an ADE model was the best fitting model, the

451 estimates of A and D were summed and entered for 'h²_ BEST' and zero was entered for
452 'c²_ BEST'. When estimates were provided separately for males and females, they were
453 entered separately; otherwise they were entered for the total sample of males and
454 females. In the case of multivariate analyses, univariate estimates were always preferred
455 to allow comparison across studies. In cases where estimates for the best fitting model
456 were not directly provided but information available in the paper indicated that the best
457 fitting model was AE (or CE, or E), we entered zero for 'c²_ BEST' and missing for 'h²_
458 BEST' (when the best fitting model was an AE model), missing for 'c²_ BEST' and zero
459 for 'h²_ BEST' (when the best fitting model was a CE model), zero for 'c²_ BEST' and
460 zero for 'h²_ BEST' (when the best fitting model was described to be an E model).

- 461 - *The total number of twin pairs as used for each entered correlation*
- 462 - *Whether or not the study was a classical twin study:* All 2,748 studies in the database
463 include MZ and DZ twins. However a classical twin study was defined as a study that
464 involved only reared-together MZ and DZ twins. From studies that included siblings,
465 extended families, adoptees, or reared apart twins only estimates based on the reared-
466 together twin sample were used for the meta-analyses. Most of these non-classical twin
467 studies did provide twin correlations for the classical twin design, and were thus
468 included in the meta-analysis for twin correlations. When A and C estimates were based
469 on extended twin designs they were excluded from the meta-analyses.
- 470 - *The method used for estimating the variance components:* We entered the statistical
471 method used for estimating the variance components, which included e.g. ANOVA,
472 Bayesian, Maximum Likelihood (ML), DeFries-Fulker regression, least squares (LS), or
473 intra-pair differences. We also listed a dichotomized version of this indicating whether

474 the method used was ‘ML or LS’ or ‘not ML or LS’ for all other methods. In the meta-
475 analyses for h^2 and c^2 estimates, only those based on ML or LS were included.

- 476 - *Whether the trait was dichotomous or continuous:* Traits measured as 0/1 as well as
477 traits measured on a quantitative scale but dichotomized before analysis, were listed as
478 dichotomous. All other traits, including ordinal traits were listed as ‘not dichotomous’
479 and treated as continuous.
- 480 - *Whether the study involved ascertainment for the trait:* When the trait under
481 investigation was the same trait that was used to select probands, the study was listed as
482 ‘ascertained’.
- 483 - *Number of concordant and discordant pairs:* In cases of dichotomous traits, the total
484 numbers of pairs for discordant and concordant affected were entered separately for each
485 zygosity. In cases of dichotomous traits that were not ascertained the number of
486 concordant unaffected pairs was also entered.
- 487 - *Prevalence:* In cases of dichotomous traits, the population prevalence, separately for MZ
488 and DZ twins when available was entered. Prevalence was based on what was provided
489 in the study or was calculated using $(2C+D)/2N$, where C is the number of concordant
490 affected pairs, D the number of discordant pairs and N is the total number of pairs in
491 non-ascertained traits.

492

493 Thus, provided availability, the statistics in **Supplementary Table S1** were obtained for each
494 trait reported on in every study. When the five basic twin correlations were available (r_{MZM} , r_{MZF} ,
495 r_{DZM} , r_{DZF} , r_{DOS}) we calculated the r_{MZall} , r_{DZSS} and r_{DZall} , using the weighted average via Fisher Z
496 transformation, and using sample size as weights. In situations where r_{MZM} (or r_{MZF}) was exactly

497 1 (or -1), we subtracted (or added in case of -1) .00001 to the correlation to ensure non-
498 problematic Fisher Z transformation. Sample sizes of MZall, DZSS and DZall were obtained by
499 summing the sample sizes of MZM and MZF, of DZM and DZF, and of DZM, DZF, and DOS
500 respectively. Estimates of h^2 and c^2 were calculated across sex as the N-weighted average across
501 the separate male and female estimates when available. For the number of concordant and
502 discordant pairs, the MZall, DZall and DZSS were calculated based on the numbers available for
503 MZM, MZF, DZM, DZF and DOS. Prevalences for pooled entries were calculated as an N-
504 weighted average.

505 **Data entry checks**

506 Studies were entered and cross-checked for obvious typos by TP and DP. After initial data-entry
507 and initial cross-checking, all data-points were manually checked a second time (DP) by looking
508 up the entered values in the original paper. In addition, automatic checks were run (DP, BB) to
509 identify typos, strange outliers, or obvious mistakes. These checks included:

- 510 - Identifying highly unlikely values (clear typos, e.g. correlation of 120)
- 511 - Testing whether the sum of h^2 and $c^2 < 100$
- 512 - Testing for strange discrepancies between estimates from the full and best fitting model
- 513 - Checking for outliers based on extreme sample size and extreme Chi-square values for
514 rejecting the null hypothesis that either $2 \times (r_{MZ} - r_{DZ})$ or $2 \times r_{DZ} - r_{MZ}$ is equal to zero.

515 **Classification of traits**

516 After data-entry, all traits were manually classified using the International Classification of
517 Functioning, Disability and Health (ICF). The ICF is the WHO's framework for health and
518 disability, and provides the conceptual basis for the definition, measurement and policy

519 formulations for health and disability. It is a universal classification of disability and health for
520 use in health and health-related sectors. ICF belongs to the World Health Organization (WHO)
521 family of international classifications, the best-known member of which is the ICD-10
522 (International Statistical Classification of Diseases and Related Health Problems). ICD-10
523 provides an etiological framework for the classification of diseases, disorders and other health
524 conditions, whereas ICF classifies functioning and disability associated with health conditions.
525 The ICD-10 and ICF are therefore complementary.

526

527 Most traits investigated in twin studies concern healthy functioning (e.g. cognitive function,
528 social attitudes, body height, personality) and could be classified according to ICF. In cases
529 where the studied traits were diseases or symptoms of disease, ICD-10 was used. Traits were
530 given two hierarchical classifications corresponding to the ICF/ICD-10 hierarchical structure,
531 using the chapter-structure (e.g. b1), and the level directly under the chapter (e.g. b110), [which](#)
532 [corresponds to the actual disease \(ICD10\) or trait \(ICF\) code.](#)

533 Six new classes at the chapter level and 17 new classes at the sub-chapter level were created to
534 accommodate traits that could not be classified under either the ICF or ICD. For the chapter level
535 the created classes were Cell, Function of DNA, Functions of the nervous system, Medication
536 effects, Mortality, and Structure of DNA. For the sub-chapter level these were All-cause
537 mortality, Cell cycle, Cell growth, Diazepam effects, Expression, Function of brain, Gene
538 expression, Height, Methylation, Mortality from heart disease, mtDNA, Physical appearance,
539 Receptor binding, Sister chromatid exchange, Structure of DNA, Telomeres, and X-inactivation.

540

541 The standardized classification schemes of the ICF and ICD-10 were queried online, via:
542 <http://apps.who.int/classifications/icfbrowser/>
543 and <http://apps.who.int/classifications/icd10/browse/2010/en> .

544

545 In addition to the two standard ICF/ICD-10 classification levels we added a general classification
546 of functional trait domains. We thus classified all traits using a three level scheme that included
547 28 broad, functional domains, 54 ICF/ICD-10 chapter level classes, and 313 subchapter level
548 classes. A small proportion of studied traits (<0.1%) could not be classified meaningfully on the
549 chapter level (2 traits) or the subchapter level (3 traits). There were 326 unique combinations
550 across the three levels of trait categorizations (**Supplementary Table S29**). All analyses were
551 conducted on all entries of each of the three levels of classification. In addition we analyzed all
552 traits together. Although this is unspecific in terms of diseases or traits, it provides a general
553 overview of the relationship between MZ and DZ twin correlations, and reveals general patterns
554 of for example sex and age differences. The most specific level is the subchapter level, which is
555 the actual ICD10 diagnosis or a similar ICF classification for normal functioning, reflecting
556 specific traits such as cleft lip, hyperkinetic disorders, or higher-level cognitive function. As
557 researchers do not necessary adhere to the ICD10 or ICF trait nomenclature, traits with the same
558 subchapter classification could have different trait names in the original study, for example for
559 Higher Level Cognitive Function the original studies included trait names total IQ score,
560 cognitive ability, intelligence, or 'g'.

561

562 **Tests for publication bias**

563 Meta-analysis relies on published results. Publication bias can occur when studies that report
564 relatively large heritability estimates or high twin correlations are more likely to be submitted
565 and/or accepted for publication than studies which report more modest effects. Such a
566 publication bias would lead to an overestimation of the true twin correlations or the true
567 heritability and environmental estimates. We used several standard statistical tools to aid in
568 identifying and quantifying possible publication bias, including inspection of funnel plots, Begg
569 and Mazumdar's test²¹, Eggers regression test²² and Rosenthal's Fail Safe N²³.

570

571 **Meta-analysis methods of twin correlations and variance components**

572 We used the DerSimonian-Laird (DSL) random-effect meta-analytical approach with correlation
573 coefficients as effect sizes, as described by Schulze²⁴ and implemented in the R package
574 'metacor'. This function transforms a correlation to its FisherZ value with corresponding
575 standard error prior to the meta-analysis. This method is preferred over conducting a meta-
576 analysis directly on the correlations because the standard error of a twin correlation is a function
577 of not only sample size but also of the correlation itself with larger correlations having a smaller
578 standard error. This can cause problems in a meta-analysis since it would lead the larger
579 correlations to appear more precise and be assigned more weight in the analysis, irrespective of
580 sample size. To avoid this problem, the DSL methods transforms correlations to the Fisher's Z
581 metric, whose standard error is determined solely by sample size. All N-weighted computations
582 were thus performed using Fisher's Z, and the results were converted back to correlations for
583 interpretation.

584

585 The random effects approach allows for heterogeneity of the true twin correlations across
586 different studies. That is, rather than assuming that there is one true level of the twin correlation,
587 the random effects model allows a distribution of true correlations. The combined effect of the
588 random effects model represents the mean of the population of true correlations. For
589 computational reasons, correlations of -1 and 1 were converted to -.99999 and .99999 prior to
590 meta-analysis. We set a threshold of at least 5 pairs of twins available per estimate and at least
591 two studies available per category. Meta-analyses were conducted for each category of all three
592 levels of classification.

593

594 We note that twin samples used in different publications were not independent. For example,
595 studies using Australian twins are predominantly based on twins from the Australian Twin
596 Registry. These studies sometimes include different subsamples, but may also include
597 completely overlapping samples used to investigate different traits. As participants are
598 anonymous, it is not possible to determine the extent of overlap in the studies included in our
599 analyses. We thus assumed independency of samples in the meta-analyses. This leads to an
600 underestimation of the variance of weighted estimates and an overestimation of their precision.
601 We expect that the dependency of study samples is lowest at the specific level of the ICD10/ICF
602 subchapter level and highest at the general functional domains.

603

604 **Meta-analysis for dichotomous, non-ascertained traits**

605 In the DSL random effects model, the standard error of a correlation is calculated based on the
606 provided N (number of twin pairs). The estimated standard errors for continuous traits are
607 correct, but for dichotomous traits the resulting standard error is incorrect. That is because twin

608 correlations for non-ascertained dichotomous traits are typically based on three categories of
609 pairs; concordant unaffected, discordant and concordant affected pairs. While the total number of
610 participating twin pairs is the sum of these three, the information that determines the twin
611 correlation and its significance is mostly derived from the latter two. For non-ascertained
612 dichotomous traits we calculated the study-specific tetrachoric twin correlation based on the
613 contingency table (i.e. the number of concordant unaffected, discordant and concordant affected
614 pairs), under the assumption that the dichotomous traits represent latent variables that follow a
615 bivariate normal distribution²⁵. We used a Maximum Likelihood estimator described by
616 Olsson²⁶, which is implemented in the R function ‘polycor’ to calculate the study specific twin
617 correlation and its standard error. As our meta-analysis required the twin correlation and sample
618 size (not standard error) as input and we wanted to be able to pool across continuous and
619 dichotomous traits, we calculated the ‘effective’ number of twin pairs based on the obtained
620 standard error. The effective number of twin pairs is defined as the number of twin pairs that
621 produces the exact same standard error within the DSL meta-analyses as the standard error
622 obtained from the contingency table.

623

624 **Meta-analysis for dichotomous, ascertained traits**

625 For ascertained traits it was not possible to calculate the twin correlations and standard errors
626 based on the contingency table as the traits include only pairs with at least one proband. Without
627 information on the number of concordant unaffected pairs, the prevalence of the affection status
628 is required to calculate a twin correlation. We used the algorithms derived from Falconer²⁷ and
629 Smith²⁸. Again, for practical purposes, calculated standard errors were transformed to an
630 effective number of twin pairs for use in the DSL meta-analysis.

631

632 **Proportion of studies consistent with specific hypotheses**

633 We estimated the proportion of studies that are consistent with $H_0: 2 \times (r_{MZ} - r_{DZ}) = 0$ ($\pi_0(h)$) and the
634 proportion of observations consistent with $H_0: 2 \times r_{DZ} - r_{MZ} = 0$ ($\pi_0(c)$), using the Jiang and Doerge
635 method²⁹, as well as the q -value method³⁰.

636 **Authorship network analysis**

637 We used the approach more fully described in Bulik-Sullivan and Sullivan³¹. Briefly, we
638 retrieved from PubMed the full Medline listing for all twin studies included in this meta-analysis
639 using NCBI “eutils”. The output was parsed to capture the names of all authors or collaborators.
640 The twin study author list was manually reviewed to resolve clear inconsistencies in the spelling
641 of the names of authors between publications. This was important so that the same author was
642 not incorrectly considered to be two people, and all were changed to the same character string as
643 to uniquely identify this researcher.

644

645 We then used GEPHI³² to construct a network in order to understand twin study publication
646 patterns. For clarity, we removed individuals who had published only one paper (i.e., we
647 required authorship on ≥ 2 papers). The substructure of the network was investigated by
648 estimating community membership modules using the Louvain method implemented in
649 GEPHI³².

650

651 **Web application**

652 The data used for this manuscript has been integrated in a web application, where user-specified
653 selections of traits can be made to apply the analyses presented in this work. The web application
654 is called **MaTCH: Meta-analysis of Twin Correlations and Heritability** and is accessible via
655 <http://match.ctglab.nl>.

656

657

- 658 21. Begg, C.B. & Mazumdar, M. Operating characteristics of a rank correlation test for
659 publication bias. *Biometrics* **50**, 1088–1101 (1994).
- 660 22. Egger, M., Davey Smith, G., Schneider, M. & Minder, C. Bias in meta-analysis detected by
661 a simple, graphical test. *BMJ* **315**, 629–634 (1997).
- 662 23. Rosenthal, R. The file drawer problem and tolerance for null results. *Psychol. Bull.* **86**, 91–
663 106 (1979).
- 664 24. Schulze, R. *Meta-analysis: A Comparison Of Approaches*. (Hogrefe & Huber, 2004).
- 665 25. Dragow, F. Polychoric and polyserial correlations. In *Encyclopedia of Statistical Sciences*
666 **7**, 68–74 (John Wiley & Sons, 1986).
- 667 26. Olsson, U. Maximum likelihood estimation of the polychoric correlation coefficient.
668 *Psychometrika* **44**, 443–460 (1979).
- 669 27. Falconer, D.S. The inheritance of liability to certain diseases, estimated from the incidence
670 among relatives. *Ann. Hum. Genet.* **29**, 51–76 (1965).
- 671 28. Smith, C. Concordance in twins: methods and interpretation. *Am. J. Hum. Genet.* **26**, 454–
672 466 (1974).

- 673 29. Jiang, H. & Doerge, R.W. Estimating the proportion of true null hypotheses for multiple
674 comparisons. *Cancer Inform.* **6**, 25–32 (2008).
- 675 30. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl.*
676 *Acad. Sci. U. S. A.* **100**, 9440–9445 (2003).
- 677 31. Bulik-Sullivan, B.K. & Sullivan, P.F. The authorship network of genome-wide association
678 studies. *Nat. Genet.* **44**, 113 (2012).
- 679 32. Blondel, V., Guillaume, J., Lambiotte, R. & Lefebvre, E. Fast unfolding of community
680 hierarchies in large networks. *J. Stat. Mech.* **10**, P10008 (2008).
- 681
- 682