Open access • Journal Article • DOI:10.1038/NG.3285

Meta-analysis of the heritability of human traits based on fifty years of twin studies

Tinca J. C. Polderman, Beben Benyamin, Christiaan de Leeuw, Patrick F. Sullivan ...+3 more authors

Institutions: VU University Amsterdam, University of Queensland, Radboud University Nijmegen, University of North Carolina at Chapel Hill ...+1 more institutions

Published on: 01 Jul 2015 - Nature Genetics (Nature Publishing Group)

Topics: Genetic correlation, Falconer's formula, Twin study and Heritability

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 Bochoven[7], 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

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

Insight into the nature of observed variation in human traits is important in medicine, psychology, social sciences and evolutionary biology. It has gained new relevance with both the ability to map genes for human traits and the availability of large, collaborative datasets to do so on an extensive and comprehensive scale. Individual differences in human traits have been studied for more than a century, yet the causes of human trait variation remain uncertain and controversial. Specifically, the partitioning of observed variability into underlying genetic and environmental sources and the relative importance of additive and non-additive genetic variation are continually debated[1-5]. Recent results from large-scale genome-wide association studies (GWAS) show that many genetic variants contribute to complex trait variation and that effect sizes are typically small[6,7]. However, the sum of the variance explained by the detected variants is much smaller than the reported heritability of the trait[4,6-10]. This 'missing heritability' has led some investigators to conclude that non-additive variation must be important[4,11]. Although the presence of gene-gene interaction has been demonstrated empirically[5,12-17], little is known about its relative contribution to observed variation[18].

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

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

RESULTS

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

We systematically retrieved published classical twin studies in which observed variation of human traits was partitioned into genetic and environmental influences. For each study we collected reported twin correlations for continuous traits and contingency tables for dichotomous traits, estimates from genetic model fitting, and study characteristics (sample size, population, age cohort and ascertainment scheme) (**Supplementary Table S1**). Investigated traits were manually classified using the chapter and subchapter levels of the International Classification of Functioning, Disability and Health (ICF) or the International Statistical Classification of Diseases and Related Health Problems (ICD-10) (**Online Methods).** The ICD10 and ICF subchapter levels refer to actual diseases (e.g. 'Atopic Dermatitis') or traits (e.g. 'Temperament and Personality Functions'). We identified 2,748 relevant twin studies, published between 1958 and 2012[20]. Half of these were published after 2004, with sample sizes per study in 2012 around

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

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

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

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

Estimates of variance components cluster in functional domains

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

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

Heterogeneity of twin correlations across sex and age

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

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

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

Falconer's equations can be used to calculate $\hat{h}^2$ and $\hat{c}^2$ based on twin correlations[18]. The 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 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 $2\times0.34-0.64=0.04$ (**Table 1, Supplementary Figure S7**). We note that the estimated $\hat{h}^2$ based on twin correlations is larger than the weighted average of reported h^2. As a consequence, the estimated $\hat{c}^2$ based on twin correlations is lower than the weighted average of the reported c^2 component. To test whether this discrepancy was due to a bias in studies reporting only twin correlations or only variance components, we conducted the meta-analysis only on studies reporting both. This yielded similar estimates with a similar discrepancy (**Supplementary Table S13)**, ruling out the explanation that twin correlations may have been reported on traits that happened to be more heritable than traits for which the estimates of variance components were reported. Through theory we show that such a discrepancy can arise when the individual studies represent a mix of traits that follow a pattern of $r_{MZ} > 2r_{DZ}$ and $r_{MZ} < 2r_{DZ}$, and where the choice of fitting a model that includes shared environment or non-additive genetic influences is based on the observed pattern of twin correlations (see **Supplementary Information, section 5.2**). More specifically, because c^2 and non-additive genetic influences cannot be estimated simultaneously from twin correlations an 'ACE' model (Additive genetic, Common environmental and Error or non-shared environmental influences) is fitted to the data if $2r_{DZ} - r_{MZ} > 0$. On the other hand, if $2r_{DZ} - r_{MZ} < 0$, an 'ADE' model, including non-additive genetic instead of common environmental influences, is selected. This leads to sampling bias in the estimate of h^2 from the full model. We show (**Supplementary Table S14**) that such per-study

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

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

There may be many causes of similarities and differences within MZ and DZ pairs and typically these are interpreted in terms of (non-) additive genetic and (non-) shared environmental influences[19]. Yet, there are essentially only two estimable and testable variance components of interest in the twin design. Therefore, inference from classical twin studies on all underlying unobserved sources of variation that lead to the resemblance between relatives is limited. However, there are two simple and parsimonious hypotheses that can be tested across traits from estimated correlation coefficients of MZ pairs (r_{MZ}) and DZ pairs (r_{DZ}). The first is that the MZ and DZ population correlations (ρ_{MZ} and ρ_{DZ}) are the same, consistent with twin resemblance being solely due to non-genetic factors. The second is a two-fold ratio of ρ_{MZ} and ρ_{DZ}, consistent with twin resemblance being solely due to additive genetic variation. An across-trait consistency with either of these hypotheses is not a proof of these simple models but would provide an extremely parsimonious model against which other experimental designs (e.g. DNA-sequence based) should be tested. For the vast majority (84%) of traits, we find that MZ correlations are larger than DZ correlations. Using the weighted estimates of r_{MZ} and r_{DZ} across all traits, we show that the average $2r_{DZ}$-r_{MZ} =.04 (SE=.01) (**Table 1**), which is very close to a two-fold difference in the MZ to DZ correlation. The proportion of single studies in which the pattern of

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

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

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

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

DISCUSSION

We have conducted a meta-analysis of virtually all twin studies published in the past 50 years, on 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.

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

References and Notes

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

11. Nelson, R.M., Pettersson, M.E. & Carlborg, Ö. A century after Fisher: time for a new paradigm in quantitative genetics. *Trends Genet. TIG* **29**, 669–676 (2013).
12. Barker, J.S. Inter-locus interactions: a review of experimental evidence. *Theor. Popul. Biol.* **16**, 323–346 (1979).
13. Cockerham, C.C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882 (1954).
14. Cockerham, C.C. in *Statistical Genetics and Plant Breeding* 53–94 (Nat. Acad. Sci. Nat. Res. Council Publ., 1963).
15. Kempthorne, O. On the covariances between relatives under selfing with general epistacy. *Proc. R. Soc. Lond. Ser. B Contain. Pap. Biol. Character R. Soc. G. B.* **144**, 100–108 (1956).
16. Crow, J.F. & Kimura, M. *An Introduction To Population Genetics Theory*. (Harper and Row, 1970).
17. Carlborg, O. & Haley, C.S. Epistasis: too often neglected in complex trait studies? *Nat. Rev. Genet.* **5**, 618–625 (2004).
18. Falconer, D.S. & Mackay, T.F.C. *Quantitative Genetics*. (Longman Group Ltd, 1996).
19. Lynch, M. & Walsch, B. *Genetics and Analysis of Quantitative Traits*. (Sinauer Associates, 1998).
20. A full reference list of al 2748 studies included in the meta-analyses is provided in **Supplementary Table S33**

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

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

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

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

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

Table 1. Weighted means of twin correlations and variance components across all human 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

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

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

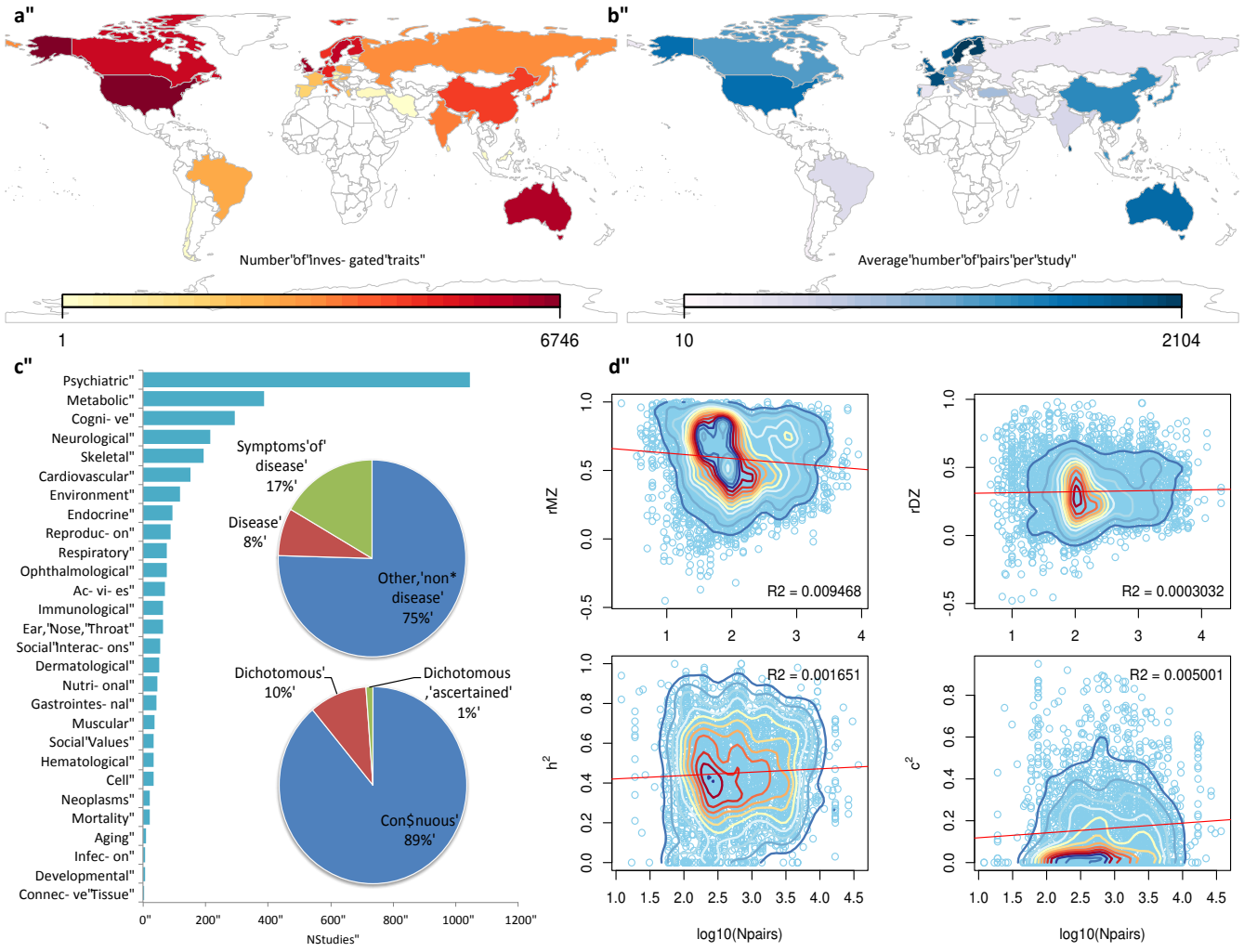
	π_0		r_{MZ}				r_{DZ}			
	NTraits	Est.	Est.	SE	NTraits	Npairs	Est.	SE	NTraits	Npairs
All traits	5185	0.69	0.64	0.0024	9568	2,563,628	0.34	0.0031	5220	2,606,252
General trait domains										
Activities	62	0.35	0.57	0.0185	118	58,227	0.34	0.0217	63	55,864
Cardiovascular	267	0.95	0.56	0.0080	380	41,669	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

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

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

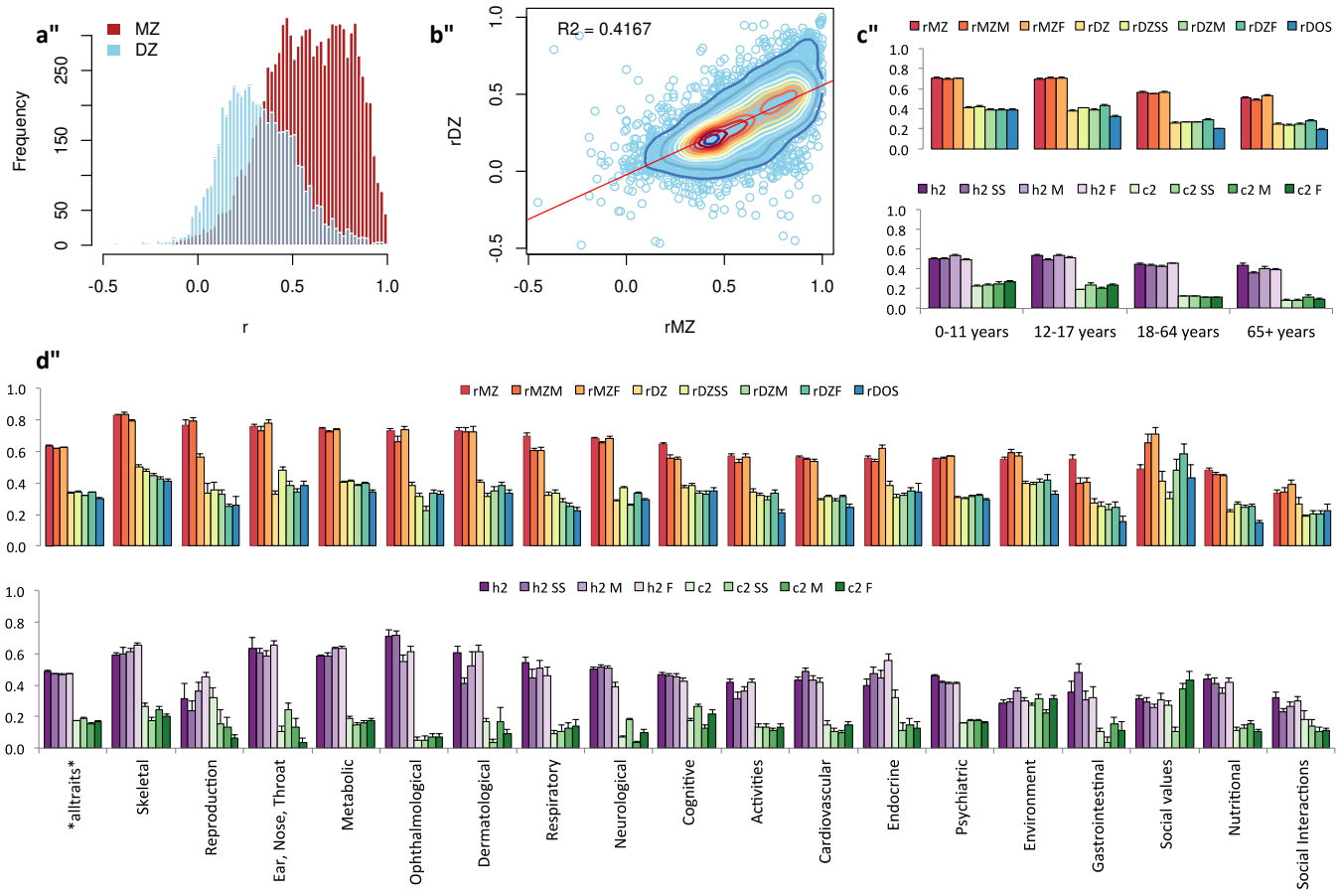
Figures and legends

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



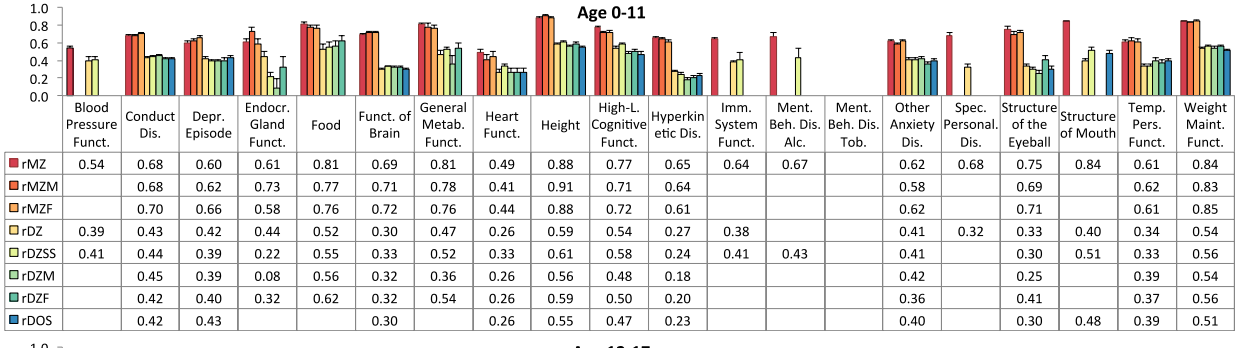
(**a**) The number of investigated traits in classical twins studies across all countries. (**b**) The average number of included pairs per study across countries. (**c**) The number of investigated traits according to functional trait domain and trait characteristic (inset). (**d**) MZ and DZ twin correlations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

Fig 2. Twin correlations and heritabilities for all human traits

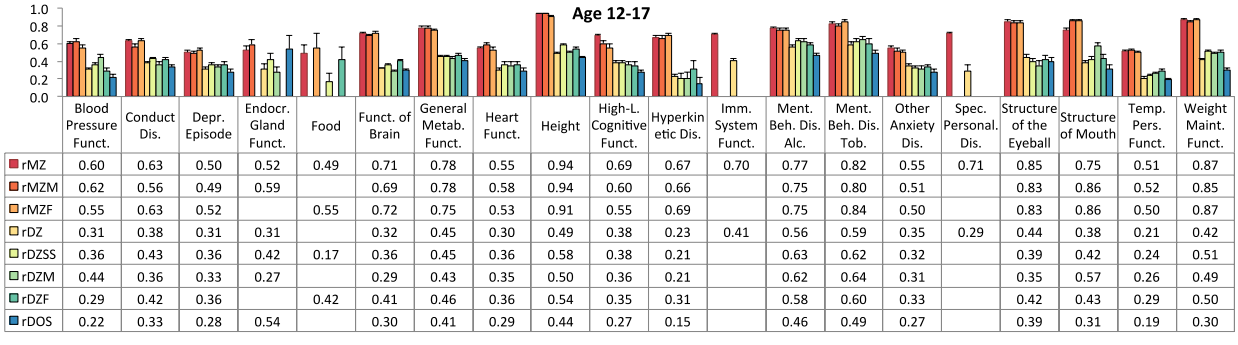


(**a**) Distribution of r_{MZ} and r_{DZ} across traits investigated in 2,748 twin studies published between 1958 and 2012. The r_{MZ} is based on 9,568 traits and 2,563,628 partly dependent pairs; r_{DZ} is based on 5,220 traits and 2,606,252 partly dependent pairs (see **Table 1**). (**b**) Relation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

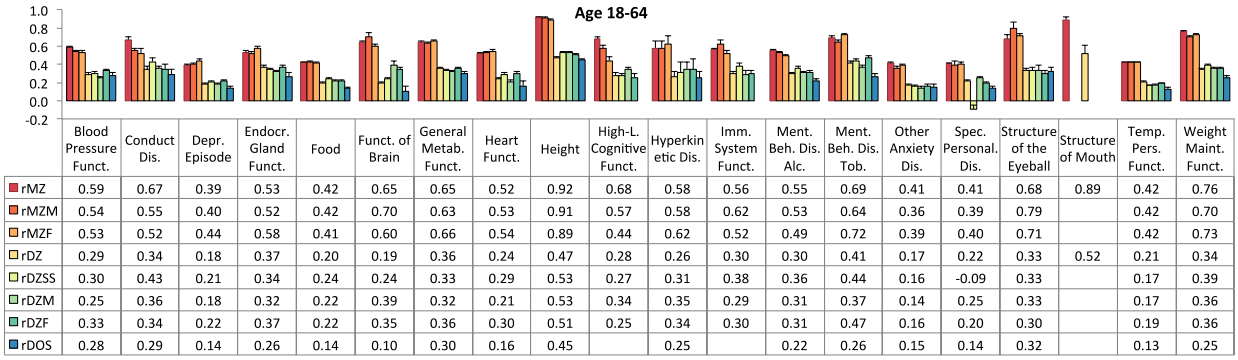
Fig. 3 Twin correlations for the top 20 most investigated specific traits across age and sex



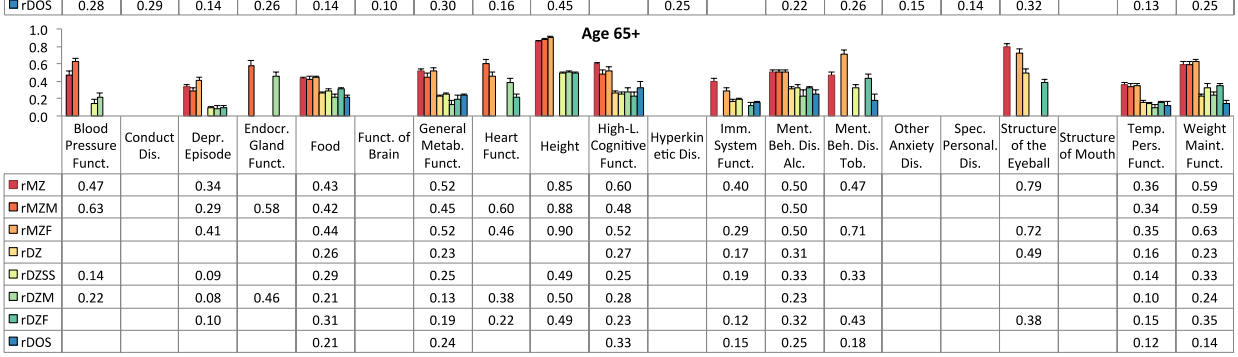
	Blood Pressure Funct.	Conduct Dis.	Depr. Episode	Endocr. Gland Funct.	Food	Funct. of Brain	General Metab. Funct.	Heart Funct.	Height	High-L. Cognitive Funct.	Hyperkinetic Dis.	Imm. System Funct.	Ment. Beh. Dis. Alc.	Ment. Beh. Dis. Tob.	Other Anxiety Dis.	Spec. Personal. Dis.	Structure of the Eyeball	Structure of Mouth	Temp. Pers. Funct.	Weight Maint. Funct.
rMZ	0.54	0.68	0.60	0.61	0.81	0.69	0.81	0.49	0.88	0.77	0.65	0.64	0.67		0.62	0.68	0.75	0.84	0.61	0.84
rMZM		0.68	0.62	0.73	0.77	0.71	0.78	0.41	0.91	0.71	0.64				0.58		0.69		0.62	0.83
rMZF		0.70	0.66	0.58	0.76	0.72	0.76	0.44	0.88	0.72	0.61				0.62		0.71		0.61	0.85
rDZ	0.39	0.43	0.42	0.44	0.52	0.30	0.47	0.26	0.59	0.54	0.27	0.38			0.41	0.32	0.33	0.40	0.34	0.54
rDZSS	0.41	0.44	0.39	0.22	0.55	0.33	0.52	0.33	0.61	0.58	0.24	0.41	0.43		0.41		0.30	0.51	0.33	0.56
rDZM		0.45	0.39	0.08	0.56	0.32	0.36	0.26	0.56	0.48	0.18				0.42		0.25		0.39	0.54
rDZF		0.42	0.40	0.32	0.62	0.32	0.54	0.26	0.59	0.50	0.20				0.36		0.41		0.37	0.56
rDOS		0.42	0.43			0.30		0.26	0.55	0.47	0.23				0.40		0.30	0.48	0.39	0.51



	Blood Pressure Funct.	Conduct Dis.	Depr. Episode	Endocr. Gland Funct.	Food	Funct. of Brain	General Metab. Funct.	Heart Funct.	Height	High-L. Cognitive Funct.	Hyperkinetic Dis.	Imm. System Funct.	Ment. Beh. Dis. Alc.	Ment. Beh. Dis. Tob.	Other Anxiety Dis.	Spec. Personal. Dis.	Structure of the Eyeball	Structure of Mouth	Temp. Pers. Funct.	Weight Maint. Funct.
rMZ	0.60	0.63	0.50	0.52	0.49	0.71	0.78	0.55	0.94	0.69	0.67	0.70	0.77	0.82	0.55	0.71	0.85	0.75	0.51	0.87
rMZM	0.62	0.56	0.49	0.59		0.69	0.78	0.58	0.94	0.60	0.66		0.75	0.80	0.51		0.83	0.86	0.52	0.85
rMZF	0.55	0.63	0.52		0.55	0.72	0.75	0.53	0.91	0.55	0.69		0.75	0.84	0.50		0.83	0.86	0.50	0.87
rDZ	0.31	0.38	0.31	0.31		0.32	0.45	0.30	0.49	0.38	0.23	0.41	0.56	0.59	0.35	0.29	0.44	0.38	0.21	0.42
rDZSS	0.36	0.43	0.36	0.42	0.17	0.36	0.45	0.36	0.58	0.38	0.21		0.63	0.62	0.32		0.39	0.42	0.24	0.51
rDZM	0.44	0.36	0.33	0.27		0.29	0.43	0.35	0.50	0.36	0.21		0.62	0.64	0.31		0.35	0.57	0.26	0.49
rDZF	0.29	0.42	0.36		0.42	0.41	0.46	0.36	0.54	0.35	0.31		0.58	0.60	0.33		0.42	0.43	0.29	0.50
rDOS	0.22	0.33	0.28	0.54		0.30	0.41	0.29	0.44	0.27	0.15		0.46	0.49	0.27		0.39	0.31	0.19	0.30



	Blood Pressure Funct.	Conduct Dis.	Depr. Episode	Endocr. Gland Funct.	Food	Funct. of Brain	General Metab. Funct.	Heart Funct.	Height	High-L. Cognitive Funct.	Hyperkinetic Dis.	Imm. System Funct.	Ment. Beh. Dis. Alc.	Ment. Beh. Dis. Tob.	Other Anxiety Dis.	Spec. Personal. Dis.	Structure of the Eyeball	Structure of Mouth	Temp. Pers. Funct.	Weight Maint. Funct.
rMZ	0.59	0.67	0.39	0.53	0.42	0.65	0.65	0.52	0.92	0.68	0.58	0.56	0.55	0.69	0.41	0.41	0.68	0.89	0.42	0.76
rMZM	0.54	0.55	0.40	0.52	0.42	0.70	0.63	0.53	0.91	0.57	0.58	0.62	0.53	0.64	0.36	0.39	0.79		0.42	0.70
rMZF	0.53	0.52	0.44	0.58	0.41	0.60	0.66	0.54	0.89	0.44	0.62	0.52	0.49	0.72	0.39	0.40	0.71		0.42	0.73
rDZ	0.29	0.34	0.18	0.37	0.20	0.19	0.36	0.24	0.47	0.28	0.26	0.30	0.30	0.41	0.17	0.22	0.33	0.52	0.21	0.34
rDZSS	0.30	0.43	0.21	0.34	0.24	0.24	0.33	0.29	0.53	0.27	0.31	0.38	0.36	0.44	0.16	-0.09	0.33		0.17	0.39
rDZM	0.25	0.36	0.18	0.32	0.22	0.39	0.32	0.21	0.53	0.34	0.35	0.29	0.31	0.37	0.14	0.25	0.33		0.17	0.36
rDZF	0.33	0.34	0.22	0.37	0.22	0.35	0.36	0.30	0.51	0.25	0.34	0.30	0.31	0.47	0.16	0.20	0.30		0.19	0.36
rDOS	0.28	0.29	0.14	0.26	0.14	0.10	0.30	0.16	0.45		0.25		0.22	0.26	0.15	0.14	0.32		0.13	0.25



	Blood Pressure Funct.	Conduct Dis.	Depr. Episode	Endocr. Gland Funct.	Food	Funct. of Brain	General Metab. Funct.	Heart Funct.	Height	High-L. Cognitive Funct.	Hyperkinetic Dis.	Imm. System Funct.	Ment. Beh. Dis. Alc.	Ment. Beh. Dis. Tob.	Other Anxiety Dis.	Spec. Personal. Dis.	Structure of the Eyeball	Structure of Mouth	Temp. Pers. Funct.	Weight Maint. Funct.
rMZ	0.47		0.34		0.43		0.52		0.85	0.60		0.40	0.50	0.47			0.79		0.36	0.59
rMZM	0.63		0.29	0.58	0.42		0.45	0.60	0.88	0.48			0.50						0.34	0.59
rMZF			0.41		0.44		0.52	0.46	0.90	0.52		0.29	0.50	0.71			0.72		0.35	0.63
rDZ					0.26		0.23			0.27		0.17	0.31				0.49		0.16	0.23
rDZSS	0.14		0.09		0.29		0.25		0.49	0.25		0.19	0.33	0.33					0.14	0.33
rDZM	0.22		0.08	0.46	0.21		0.13	0.38	0.50	0.28			0.23						0.10	0.24
rDZF			0.10		0.31		0.19	0.22	0.49	0.23		0.12	0.32	0.43			0.38		0.15	0.35
rDOS					0.21		0.24			0.33		0.15	0.25	0.18					0.12	0.14

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

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

Online Methods

Identifying relevant studies

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

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

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

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

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

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

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

Primary information obtained from each study

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

- *The country of origin of the study population:* We used standard ISO country names and where possible data entry was done separately for each country investigated in the study. Note that in studies where pooled estimates across different countries were provided, yet sample sizes were available for each country separately, we entered information separately for each country.
- *The age group of the study population:* The study population was classified into four age cohorts based on the average age of the population included in the study, as: age > 0 and < 12; age >= 12 and < 18; age >= 18 and < 65; and age >= 65.
- *The MZ and DZ twin correlations:* Twin correlations were entered as provided in the study, and could be calculated as intra-class, Pearson, polychoric or tetrachoric

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

- *The estimates of the heritability (h^2) and shared environmental component (c^2), under the full ACE (or ADE) model:* We entered 'h^2_FULL' and 'c^2_FULL', based on estimates under the full ACE (including additive genetic, shared and non-shared environmental influences) or ADE (including additive and non-additive genetic and non-shared environmental influences) model. When an ACE model was fitted, the estimate for A was entered in 'h^2_FULL' and the estimated for C in 'c^2_FULL'. When an ADE model was fitted, the estimates of A and D were summed and entered for 'h^2_FULL' and zero was entered for 'c^2_FULL'. When estimates were provided separately for males and females, they were entered separately; otherwise they were entered for the total sample of males and females. In the case of multivariate analyses, univariate estimates were always preferred to allow comparison across studies.
- *The estimates of the heritability (h^2) and shared environmental component (c^2), under the best fitting ACE (or ADE) model:* We entered 'h^2_BEST' and 'c^2_BEST', based on estimates under the best fitting ACE or ADE model as provided in the study. When an ACE model was the best fitting model, the estimate for A was entered in 'h^2_BEST' and the estimate for C in 'c^2_ BEST'. When an ADE model was the best fitting model, the

estimates of A and D were summed and entered for 'h^2_ BEST' and zero was entered for 'c^2_ BEST'. When estimates were provided separately for males and females, they were entered separately; otherwise they were entered for the total sample of males and females. In the case of multivariate analyses, univariate estimates were always preferred to allow comparison across studies. In cases where estimates for the best fitting model were not directly provided but information available in the paper indicated that the best fitting model was AE (or CE, or E), we entered zero for 'c^2_ BEST' and missing for 'h^2_ BEST' (when the best fitting model was an AE model), missing for 'c^2_ BEST' and zero for 'h^2_ BEST' (when the best fitting model was a CE model), zero for 'c^2_ BEST' and zero for 'h^2_ BEST' (when the best fitting model was described to be an E model).

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

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

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

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

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

Data entry checks

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

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

Classification of traits

After data-entry, all traits were manually classified using the International Classification of Functioning, Disability and Health (ICF). The ICF is the WHO's framework for health and 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

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

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

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 test[21], Eggers regression test[22] and Rosenthal's Fail Safe N[23].
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 Schulze[24] 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

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

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

Meta-analysis for dichotomous, non-ascertained traits

In the DSL random effects model, the standard error of a correlation is calculated based on the provided N (number of twin pairs). The estimated standard errors for continuous traits are 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 distribution[25]. We used a Maximum Likelihood estimator described by
616 Olsson[26], 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 Falconer[27] and
629 Smith[28]. 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:$2\times(r_{MZ}-r_{DZ})=0$ (π_0(h)) and the
634 proportion of observations consistent with H_0:$2\times r_{DZ}-r_{MZ}=0$ (π_0(c)), using the Jiang and Doerge
635 method[29], as well as the *q*-value method[30].

636 **Authorship network analysis**

637 We used the approach more fully described in Bulik-Sullivan and Sullivan[31]. 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 GEPHI[32] 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 GEPHI[32].

650

Web application

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

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