EJN European Journal of Neuroscience

FENS Federation of European Neuroscience Societies

European Journal of Neuroscience, Vol. 47, pp. 109–114, 2018 doi:10.1111/ejn.13801

EDITORIAL
Neurogenomics – towards a more rigorous science

Kevin J. Mitchell
Institutes of Genetics and Neuroscience, Trinity College Dublin, Dublin 2, Ireland

Overview

The field of neurogenomics is coming of age, but not without some teething problems. The aim of this field is to understand the genetic basis of differences in brain structure and function, which in turn underlie differences in behaviour, cognition, perception, mood and other psychological faculties, as well as susceptibility to psychiatric disorders. Both imaging and genomic technologies are now being applied on a very large scale, greatly boosting the potential power of this combined approach. However, the field is lagging behind some others in the degree of rigour and quality control that is demanded. Past practices have yielded a literature hopelessly polluted with spurious findings. To make real progress, the field will have to learn from these mistakes and adopt more rigorous standards. In addition, our emerging understanding of the genetic architecture of psychological traits or psychiatric disorders has important implications for the design and interpretation of imaging and neurogenomics experiments.

Introduction

Twin and family studies have shown that most behavioural and cognitive phenotypes are at least moderately heritable. This includes personality traits, which can be assessed in a pseudoquantitative fashion using questionnaires, or cognitive traits, such as intelligence, memory or a wide range of more specific faculties, which can be assessed by psychometric tests. It even extends to observed occurrences of specific behaviours or life outcomes, such as educational attainment, physical violence or divorce. For all of these phenotypes, a substantial proportion of the variance across the population is attributable to genetic variation (Polderman *et al.*, 2015). The same is true for psychiatric disorders, most of which also show high levels of heritability (Polderman *et al.*, 2015). However, in all of these cases, the measurement of the phenotype itself is a limiting factor – most psychological tests and even psychiatric diagnoses are somewhat fuzzy. One of the goals of neurogenomics is to move to a level deeper – to define brain-based phenotypes that are more precisely measurable and that may also reveal the biological mechanisms underlying variation in behavioural or cognitive traits.

The application of neuroimaging technologies in twins and family members, and also across only distantly related individuals, has shown that a wide range of brain-based parameters is also quite highly heritable (Jansen *et al.*, 2015; Toro *et al.*, 2015). On a structural level, these include overall brain size, grey matter volume, white matter volume, sizes of various brain regions, fractional anisotropy or other diffusion-based parameters. The same is true for functional parameters, both in the response of particular brain regions during specific tasks, or in more global resting state measures. This can also be extended to global measures of connectivity or other network parameters such as small-worldness, modularity or efficiency, which all also show at least moderate heritability (Fornito & Bullmore, 2012). The hope in neurogenomics is now to identify specific genetic variants contributing to these differences across the normal range or to alterations in these parameters in disease. This should in turn provide entry points to further elucidate the molecular and neural mechanisms underlying the various cognitive and behavioural faculties. However, the methods used to date in this field have been plagued by statistical deficiencies and less than rigorous research practices. In addition, assumptions about the genetic basis of these kinds of phenotypes have turned out be naïve.

Methodological issues

Much of the research in this field has relied on candidate gene association studies. These studies analyse the relative frequency of common genetic variants in some particular gene or genes of interest, across people with varying levels of some brain-based phenotype. Throughout the human genome, there are millions of sites where the DNA sequence comes in two versions – it might be an 'A' in 30% of people and a 'T' in 70%. (Really, this should read 30% of *chromosomes*, as we each carry two copies, so in the example above, some people would have an AA genotype, others AT and others TT). An association study simply asks whether one or other of the versions of these so-called single nucleotide polymorphisms (SNPs), or one of the genotypes, differs in frequency between two phenotypically distinct groups or along a phenotypic continuum. If it does, the inference is that either that genetic variant itself, or another variant nearby on the chromosome, is causally contributing to the phenotypic difference.

In the 2000's, an international effort mapped the common variation across the genome, thus allowing researchers to select particular SNPs in their gene of interest and test association with their phenotypes of interest (Consortium, International HapMap, 2003). The nice thing about this is that the methods for genotyping people – determining which versions of the SNP they carry – are cheap and straightforward. So, if you have a sample of people with some imaging phenotypes, performing an association study is relatively trivial. In addition, an explicit argument was made that the sample sizes required to detect a genetic association should be smaller for structural and functional brain-based phenotypes than for ones measured in cognitive or behavioural tests. This was based on the idea that brain-based phenotypes would be 'closer to the action of the genes' than behavioural ones and that the genetic effects of common variants would therefore be much larger in size (Gottesman &

Correspondence: K. J. Mitchell, as above.
E-mail: Kevin.Mitchell@tcd.ie

Gould, 2003; Glahn *et al.*, 2007). Despite early reports of such large effects, this has not held true. The genetics of these phenotypes do not appear to be any simpler than that of psychological traits (Glahn *et al.*, 2014; Bearden & Thompson, 2017).

Candidate gene association studies are fine in principle, but in practice have been beset by a host of methodological problems (Flint & Munafo, 2013; Dumas-Mallet *et al.*, 2017). These apply whatever the phenotype, but for neuroimaging experiments, the problems are compounded due to the increased number of possible phenotypes to analyse (Ioannidis, 2005; Button *et al.*, 2013). These problems include the following: (i) small sample sizes; (ii) failure to correct for testing of multiple SNPs; (iii) lack of pre-defined specific hypotheses; (iv) excess researcher degrees of freedom in deriving neuroimaging parameters to use as phenotypes; (v) lack of a replication sample; and (vi) publication bias favouring reporting of positive findings. If you look at enough SNPs or genotypes, and define enough different neuroimaging phenotypes, where any difference in any direction in any parameter will do, you are certain to see something that appears statistically significant in any given sample of limited size, or at least in many such samples. Selectively reporting only the positive findings will make them seem much less spurious than they actually are.

These practices, which have been standard operating procedure in the field for many years, have produced a body of literature that is dominated by false positives. There are much celebrated and familiar examples of variants in genes like COMT (catechol-O-methyltransferase), BDNF (brain-derived neurotrophic factor), MAOA (monoamine oxidase-A) and SLC6A4 (which encodes the serotonin transporter), which have been associated with a huge array of structural and functional brain phenotypes, (reviewed in (Gelernter, 2014; Hashimoto *et al.*, 2015; Bogdan *et al.*, 2017)), many with very large reported effect sizes (Egan *et al.*, 2001, 2003; Hariri *et al.*, 2002; Buckholtz *et al.*, 2008; Gelernter, 2014). And there are scores of other similar reported associations in less well-studied genes. Unfortunately, these reports have not stood up to replication (Bogdan *et al.*, 2017). The same is true of candidate gene by environment associations (Duncan & Keller, 2011).

In general, candidate gene association studies have proven to be unreliable (Ioannidis *et al.*, 2011), and the human genetics field has largely stopped doing them. Regrettably, these lessons have not all been taken on board by the neurogenomics community. Candidate gene association studies with imaging phenotypes continue to be carried out with all of the methodological problems referred to above. Clearly they are still being positively reviewed, both for funding applications and for publication, as they continue to appear in the literature on a regular basis, including these examples just from the last several months: (Bruce *et al.*, 2017; Dalvie *et al.*, 2017; Gonzalez *et al.*, 2017; Jasinska *et al.*, 2017; Lubeiro *et al.*, 2017; Mallas *et al.*, 2017; Swartz *et al.*, 2017). This is despite the very solid empirical finding that most, perhaps all, such reported results will not generalise beyond the sample studied.

These problems pertain to all candidate gene association studies, not just those using neuroimaging phenotypes. Thankfully, they were recognised by the human genetics community, and steps were taken to remedy them. Foremost among these was the development of technology to enable genomewide association studies (GWASs) (Visscher *et al.*, 2017). Rather than testing SNPs only in a given gene, GWASs allow researchers to look for frequency differences in hundreds of thousands of SNPs across the entire genome. Because this incurs a huge multiple testing burden, it requires massive samples to achieve genomewide statistical significance for any given SNP. This has been achieved through the development of large international consortia, which have pooled samples to reach the requisite critical mass.

These consortia have developed rigorous statistical methods to control for multiple testing and also demanded inclusion of separate replication samples in the initial design of each study to ensure that any primary findings are robust. Because GWASs report findings for all SNPs, they also get around the pernicious problem of publication bias. GWASs have now successfully identified hundreds of SNPs associated with all kinds of traits (Visscher *et al.*, 2017), including psychological ones such as intelligence (Direk *et al.*, 2017) and neuroticism (Luciano, 2017), as well as ones associated with risk of psychiatric disorders, such as schizophrenia (Consortium, Schizophrenia Working Group of the International Psychiatric Genetics, 2014) and depression (Direk *et al.*, 2017).

For neuroimaging-based studies, this is more challenging, as it is much more time-consuming and expensive to collect large samples with neuroimaging phenotypes and also more difficult to standardise phenotyping across collection centres using different equipment and protocols. Nevertheless, several large neuroimaging consortia have been set-up, including ADNI (Shen *et al.*, 2010) ENIGMA (Bearden & Thompson, 2017), IMAGEN (Schumann *et al.*, 2010), CHARGE (Psaty *et al.*, 2009) and the UK Biobank (Miller *et al.*, 2016), for example, and the sample sizes in these databases continue to grow. These approaches have begun to yield positive results (Thompson *et al.*, 2017), with the identification of a small number of common variants contributing to variation in hippocampal volume, for example (Bis *et al.*, 2012; Stein *et al.*, 2012; Elliott, 2017; Hibar *et al.*, 2017) and a very recent, though still preliminary, report of a large number of associations with diverse imaging phenotypes (Elliott, 2017).

Even at this early stage, there are a number of general conclusions that can be drawn from these studies. First, none of the reported candidate gene effects on brain anatomy have replicated in these larger GWASs (Bogdan *et al.*, 2017; Jahanshad, 2017). Reported effects on size of various brain regions, such as that of a variant in BDNF on the hippocampus, for example (Egan *et al.*, 2003), have not shown up in much more highly powered, unbiased whole-genome analyses (Bis *et al.*, 2012; Stein *et al.*, 2012; Elliott, 2017; Hibar *et al.*, 2017). Second, there are no common variants that have even modest individual effects on brain structure. It is not just that they have not been found – these studies were very well powered to detect them if they existed. The negative result shows quite conclusively that they do not. Identifying effects of common variants will thus require much larger samples.

Finally, a number of techniques have been developed to use GWAS statistics to estimate the total contribution of common variants to the heritability of the trait being studied (Vinkhuyzen *et al.*, 2013). Or, more accurately, the amount of variance signal that can be *tagged by* common variation, as the methods only use common variants to index distant relatedness and compare this with phenotypic similarity. As distant relatives actually share only small blocks of the genome by descent, this signal may be driven by either rare or common variants within such blocks. The application of these techniques to brain-based phenotypes suggests that 40-50% of the heritability of a variety of structural or functional measures can be tagged in this way by common variants (Dickie *et al.*, 2014; Chen *et al.*, 2015; Toro *et al.*, 2015; Elliott, 2017; Thompson *et al.*, 2017). This suggests many more exist to be found, but these estimates also put an upper bound on the collective effects of all common variants. They therefore provide strong evidence that much of the variation in brain structure and function will be due to rare genetic variants that are not captured by GWAS.

Theoretical considerations

One of the rationales for studying brain-based phenotypes is that they can act as 'endophenotypes' for psychiatric disorders (Gottesman & Gould, 2003; Glahn *et al.*, 2007). The hope was that the genetics of, say, working memory and of the neural correlates that underlie it, would be simpler than the genetics of a condition like schizophrenia, which affects this faculty. If many genetic variants influence risk of schizophrenia, and working memory is only one of the things affected, then maybe a smaller subset of the variants would specifically affect working memory (and likewise for all the other symptoms). If that were true, then common genetic variants with small effects on schizophrenia risk might have larger and more direct effects on the neural substrates of working memory.

That this turns out not to be the case should not be surprising, for two reasons. First, there is no reason to think that the genetic architecture of neural function should be modular. Most genetic variants that affect one brain region or function will also affect many others and most of their effects are highly indirect and emergent. There are no genes 'for working memory' or for its supposed neural correlates, such as hippocampal–prefrontal neuronal synchrony in the gamma frequency range, any more than there are genes 'for schizophrenia'. There are certainly genetic variants that can affect these things, but they are not specific and the genes involved are not dedicated to those functions. Brain-based phenotypes are affected by hundreds of genetic variants in any given person, in the same complex and indirect ways that behavioural or cognitive traits are. When we look inside the big black box, we should not expect to see lots of smaller black boxes – it is a mess in there. Second, we should not expect common variants to have large effects, for evolutionary reasons. Most large effects are bad, because it is simply vastly easier to break an already highly optimised system than to improve it with random tinkering. Most variants that cause large effects are consequently selected against, and thus never become common (Keller & Miller, 2006).

Rare variants, on the other hand, can have much larger effects. This is particularly relevant to psychiatric disorders, where rare mutations in hundreds of different genes have been found to be responsible for a sizeable fraction of cases of intellectual disability, autism, schizophrenia, epilepsy and other conditions (Mitchell, 2015), with more being reported on a weekly basis. In addition, there is very strong overlap in genetic risk across what were previously thought of as distinct psychiatric disorders – mutations that predispose to one condition almost always predispose to many (Moreno-De-Luca *et al.*, 2013; Mitchell, 2015). These findings challenge the conception of psychiatric categories as natural kinds. They are more accurately thought of as a set of psychopathological and pathophysiological states that the brain can end up in, in response to any of a very large number of different genetic insults. Diagnostic categories like autism or schizophrenia are thus, like intellectual disability, umbrella terms for a hugely heterogeneous group of genetic disorders. This has important implications for the design of neuroimaging studies looking for phenotypes associated with diagnostic categories.

Recommendations

Increase sample sizes

This seems the most obvious means of reducing the occurrence of spurious results due to random sampling error, especially for group comparisons. It will likely require collaborations and the establishment of large consortia, as achieved in many areas of disease genetics. It is important to note, however, that by itself, increasing sample size will not solve the problem of false positives. It will simply lead to greater power to attach statistical significance to effects of smaller and smaller size.

Raise the P*-value threshold?*

Recent suggestions of lowering the *P*-value significance threshold from <0.05 to <0.005 are welcome (Benjamin *et al.*, 2017). This should certainly reduce the flood of statistical blips that are published as real findings. On the other hand, if you have a well-defined hypothesis and a well-designed experiment to test it, $P < 0.05$ may be perfectly adequate. By contrast, if you are doing exploratory analyses, then no *P*-value threshold is really appropriate, because inferential statistics should not be applied to exploratory data under most circumstances.

Distinguish exploratory analyses from hypothesis testing

Many neurogenomics studies have only the vaguest possible hypothesis – namely, that something, somewhere in the brain will differ in some way between people carrying different versions of some genetic variant (often any one of many such variants). You can set a false discovery rate, using a variety of statistical methods (more on that below), but it is simply not appropriate to perform null hypothesis testing on the resulting data. You cannot test a hypothesis on the same data that suggested it.

Richard Feynman once famously started a lecture with a story about reasoning from known facts back to possible causes. '*You know, the most amazing thing happened to me tonight… I saw a car with the license plate ARW 357. Can you imagine? Of all the millions of license plates in the state, what was the chance that I would see that particular one tonight? Amazing!'* As he later articulated, in more precise terms: '*To report a significant result and reject the null in favor of an alternative hypothesis is meaningless unless the alternative hypothesis has been stated before the data was obtained'*. There is nothing wrong with exploratory analyses – this is how we generate hypotheses. But those hypotheses can only be confirmed in independent experiments.

Define false discovery rates empirically

If I compare a thousand different imaging measurements between two groups of random people, what will the frequency be of observing some statistically significant differences, given a sample size of 20, or 50, or 1000, or 5000? Given the public availability of large imaging data sets, it should be possible to directly measure this false discovery rate using real data from random sets of people, for any given set of neuroimaging parameters, as opposed to relying on the variety of statistical methods for estimating it from the same sample you are testing. Such false discovery rates are clearly not well calibrated, judging by the failure to replicate most results reported as surpassing that threshold. Empirically determining false discovery rates should allow researchers to more rigorously define which findings from exploratory analyses are truly surprising and worthy of independent testing.

Restrict researcher degrees of freedom

In most neuroimaging experiments, there are dozens of different parameters that can be played with to look for an effect and scores of different measures that can be treated as phenotypes. You can look at global differences, or by region of interest, or voxel-wise, with varying cluster size and thresholds of significance. You can

look at volume, thickness, surface area, fractional anisotropy, mean diffusivity, the size of specific axonal tracts. You can define any number of task-based or resting state functional parameters. And you can extract global measures of connectivity or other network parameters that act as meta-phenotypes. If these parameters or choices of phenotype are tweaked with the data in hand, based on hints of an effect, it is almost inevitable that some spurious difference will emerge (Silver *et al.*, 2011; Bennett *et al.*, 2012; Button *et al.*, 2013). Every additional parameter change, and every covariate thrown into the mix, exponentially increases the multiple testing burden, though these are rarely corrected for. Ideally, the details of the analytical pipeline should be pre-defined and, if possible, pre-registered. If it really is an exploratory analysis, this should simply be acknowledged and the inferences appropriately limited.

Replicate, replicate, replicate

Other areas of human genetics have recognised the need for replication and adopted it as standard practice. There is no reason why imaging phenotypes should be different – in fact, you could argue it should be especially required for imaging phenotypes, given the huge number of degrees of freedom in defining them. The results of exploratory analyses are simply things that *could* differ between your groups of interest; they are not strong evidence that they actually generalise beyond the analysed sample, and most of them will not. This is true no matter the sample size, if the effects are small. Any suspected phenotypic differences should be tested specifically in an independent replication sample, as a condition of publication.

Some researchers have objected to calls for larger samples and for independent replication samples by saying it is simply impractical or too expensive for most research groups to collect them. The argument is that such groups should be allowed to continue doing smaller studies, that such studies should continue to be funded and published, and that this will provide essential training for graduate students who otherwise might have nothing to do. None of these arguments is compelling. Underpowered, exploratory studies with high degrees of freedom and without replication samples simply generate noise, waste everyone's time and resources, and pollute the literature with false positives. They are worse than doing nothing.

Polygenic scores

Individual common variants are likely to have such small effects that it may require enormous samples to identify them individually. However, it is possible to generate polygenic scores based on the highest-ranking sets of SNPs in a GWAS, which can be used to look at the collective effect of many SNPs that may be involved in the trait (Dima & Breen, 2015). Such scores can be treated as a genetic variant that can be tested for association with all kinds of brain phenotypes, and which presumably could show much greater effect sizes. This approach has potential but is limited in two important ways. First, most such polygenic scores explain only a small percentage of the variance in the primary phenotype that they index. For example, the most recent GWAS for intelligence identifies several hundred significantly associated SNPs (Direk *et al.*, 2017). However, a polygenic score that incorporates their collective effects explains only 4% of the variance in intelligence in another sample. Polygenic scores for schizophrenia also explain only 3–7% of variance (Consortium, Schizophrenia Working Group of the International Psychiatric Genetics, 2014). Finding that such a score also tracks some brain phenotype would be interesting, but its interpretation should be tempered by the marginal effects at play. Second, using polygenic scores as the genotype in question still leaves all the other problems of excess researcher degrees of freedom and uncorrected multiple testing that can arise in exploratory neuroimaging analyses. Any exploratory findings should still be replicated in an independent sample.

Genotype first

The genetic and clinical heterogeneity of psychiatric categories will make it almost impossible to extract meaningful information from analyses of groups of patients with 'autism' or 'schizophrenia' or 'depression'. A hundred people with a diagnosis of autism may have a hundred different underlying causes. This heterogeneity is a likely cause of the failure of the extremely extensive psychiatric neuroimaging literature to discover a single consistent imaging biomarker that is specific for any diagnostic category (Sprooten *et al.*, 2017). An alternative approach, made possible by the large-scale sequencing programs now underway in many places, is to identify smaller subsets of patients with defined genetic conditions (Mefford, 2009; Stessman *et al.*, 2014). This list used to be limited to things like Fragile X syndrome, Rett syndrome, 22q11 deletion syndrome and a few others, but there are now hundreds of such conditions identified. Of course, the sample sizes will be much smaller, but the relative homogeneity may enable identification of larger differences in brain structure or function that characterise these rare conditions, though replication will still be essential (Mahmood *et al.*, 2010; Consortium, Simons VIP, 2012; Green *et al.*, 2015; Schmitt *et al.*, 2015; Ulfarsson *et al.*, 2017). Such differences would otherwise be obscured by group comparisons that bundle subjects together based on psychiatric diagnoses alone.

Conclusion

The revolution in genomics and the ever-increasing sample sizes of neuroimaging databases present a golden opportunity to uncover the genetic basis of variation in brain structure and function. But realising this goal will require more rigorous approaches that have typically been employed to date, and ones that are better grounded in our understanding of the complex genetic architecture of these traits.

References

Bearden, C.E. & Thompson, P.M. (2017) Emerging global initiatives in neurogenetics: the enhancing neuroimaging genetics through meta-analysis (ENIGMA) consortium. *Neuron*, **94**, 232–236.

Benjamin, D.J., Berger, J., Johannesson, M., Nosek, B.A., Wagenmakers, E.-J., Berk, R. & Johnson, V. (2017) Redefine statistical significance. *PsyArXiv. July 22. psyarxiv.com/mky9j.*

Bennett, C.M., Baird, A.A., Miller, M.B. & Wolford, G.L. (2012) Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for multiple comparisons correction. *J Serendipitous Unexpect Res*, **1**, 1–5.

Bis, J.C., DeCarli, C., Smith, A.V., van der Lijn, F., Crivello, F., Fornage, M., Debette, S., Shulman, J.M. *et al.* (2012) Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat. Genet.*, **44**, 545–551.

Bogdan, R., Salmeron, B.J., Carey, C.E., Agrawal, A., Calhoun, V.D., Garavan, H., Hariri, A.R., Heinz, A. *et al.* (2017) Imaging genetics and genomics in psychiatry: a critical review of progress and potential. *Biol. Psychiat.*, **82**, 165–175.

Bruce, H.A., Kochunov, P., Paciga, S.A., Hyde, C.L., Chen, X., Xie, Z., Zhang, B., Xi, H.S. *et al.* (2017) Potassium channel gene associations with joint processing speed and white matter impairments in schizophrenia. *Genes Brain Behav.*, **16**, 515–521.

Buckholtz, J.W., Callicott, J.H., Kolachana, B., Hariri, A.R., Goldberg, T.E., Genderson, M., Egan, M.F., Mattay, V.S. *et al.* (2008) Genetic variation in MAOA modulates ventromedial prefrontal circuitry mediating individual differences in human personality. *Mol. Psychiatry*, **13**, 313–324.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S. & Munafo, M.R. (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.*, **14**, 365–376.

Chen, C.H., Peng, Q., Schork, A.J., Lo, M.T., Fan, C.C., Wang, Y., Desikan, R.S., Bettella, F. *et al.* (2015) Large-scale genomics unveil polygenic architecture of human cortical surface area. *Nat. Commun.*, **6**, 7549.

Consortium, International HapMap (2003) The international HapMap project. *Nature*, **426**, 789–796.

Consortium, Schizophrenia Working Group of the International Psychiatric Genetics (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421–427.

Consortium, Simons VIP (2012) Simons variation in individuals project (Simons VIP): a genetics-first approach to studying autism spectrum and related neurodevelopmental disorders. *Neuron*, **73**, 1063–1067.

Dalvie, S., Brooks, S.J., Cardenas, V., Fein, G., Ramesar, R. & Stein, D.J. (2017) Genetic variation within GRIN2B in adolescents with alcohol use disorder may be associated with larger left posterior cingulate cortex volume. *Acta Neuropsychiatr*, **29**, 252–258.

Dickie, E.W., Tahmasebi, A., French, L., Kovacevic, N., Banaschewski, T., Barker, G.J., Bokde, A., Buchel, C. *et al.* (2014) Global genetic variations predict brain response to faces. *PLoS Genet.*, **10**, e1004523.

Dima, D. & Breen, G. (2015) Polygenic risk scores in imaging genetics: usefulness and applications. *J Psychopharmacol*, **29**, 867–871.

Direk, N., Williams, S., Smith, J.A., Ripke, S., Air, T., Amare, A.T., Amin, N., Baune, B.T. *et al.* (2017) An analysis of two genome-wide association meta-analyses identifies a new locus for broad depression phenotype. *Biol. Psychiat.*, **82**, 322–329.

Dumas-Mallet, E., Button, K.S., Boraud, T., Gonon, F. & Munafo, M.R. (2017) Low statistical power in biomedical science: a review of three human research domains. *R Soc Open Sci*, **4**, 160254.

Duncan, L.E. & Keller, M.C. (2011) A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am. J. Psychiatry*, **168**, 1041–1049.

Egan, M.F., Goldberg, T.E., Kolachana, B.S., Callicott, J.H., Mazzanti, C.M., Straub, R.E., Goldman, D. & Weinberger, D.R. (2001) Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc. Natl. Acad. Sci. USA*, **98**, 6917–6922.

Egan, M.F., Kojima, M., Callicott, J.H., Goldberg, T.E., Kolachana, B.S., Bertolino, A., Zaitsev, E., Gold, B. *et al.* (2003) The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell*, **112**, 257–269.

Elliott, L. (2017) The genetic basis of human brain structure and function: 1,262 genome-wide associations found from 3,144 GWAS of multimodal brain imaging phenotypes from 9,707 UK Biobank participants. *bioRxiv* 178806; https://doi.org/10.1101/178806, [Epub ahead of print].

Flint, J. & Munafo, M.R. (2013) Candidate and non-candidate genes in behavior genetics. *Curr. Opin. Neurobiol.*, **23**, 57–61.

Fornito, A. & Bullmore, E.T. (2012) Connectomic intermediate phenotypes for psychiatric disorders. *Front. Psychiatry.*, **3**, 32.

Gelernter, J. (2014) SLC6A4 polymorphism, population genetics, and psychiatric traits. *Hum. Genet.*, **133**, 459–461.

Glahn, D.C., Thompson, P.M. & Blangero, J. (2007) Neuroimaging endophenotypes: strategies for finding genes influencing brain structure and function. *Hum. Brain Mapp.*, **28**, 488–501.

Glahn, D.C., Knowles, E.E., McKay, D.R., Sprooten, E., Raventos, H., Blangero, J., Gottesman, I.I. & Almasy, L. (2014) Arguments for the sake of endophenotypes: examining common misconceptions about the use of endophenotypes in psychiatric genetics. *Am J Med Genetics*, **165B**, 122–130.

Gonzalez, D.A., Jia, T., Pinzon, J.H., Acevedo, S.F., Ojelade, S.A., Xu, B., Tay, N., Desrivieres, S. *et al.* (2017) The Arf6 activator Efa6/PSD3 confers regional specificity and modulates ethanol consumption in Drosophila and humans. *Mol. Psychiatry*. https://doi.org/10.1038/mp.2017.112, [Epub ahead of print].

Gottesman, I.I. & Gould, T.D. (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry*, **160**, 636–645.

Green, T., Barnea-Goraly, N., Raman, M., Hall, S.S., Lightbody, A.A., Bruno, J.L., Quintin, E.M. & Reiss, A.L. (2015) Specific effect of the fragile-X mental retardation-1 gene (FMR1) on white matter microstructure. *Brit J Psychiatry*, **207**, 143–148.

Hariri, A.R., Mattay, V.S., Tessitore, A., Kolachana, B., Fera, F., Goldman, D., Egan, M.F. & Weinberger, D.R. (2002) Serotonin transporter genetic variation and the response of the human amygdala. *Science*, **297**, 400–403.

Hashimoto, R., Ohi, K., Yamamori, H., Yasuda, Y., Fujimoto, M., Umeda-Yano, S., Watanabe, Y., Fukunaga, M. *et al.* (2015) Imaging genetics and psychiatric disorders. *Curr. Mol. Med.*, **15**, 168–175.

Hibar, D.P., Adams, H.H., Jahanshad, N., Chauhan, G., Stein, J.L., Hofer, E., Renteria, M.E., Bis, J.C. *et al.* (2017) Novel genetic loci associated with hippocampal volume. *Nat. Commun.*, **8**, 13624.

Ioannidis, J.P. (2005) Why most published research findings are false. *PLoS Med.*, **2**, e124.

Ioannidis, J.P., Tarone, R. & McLaughlin, J.K. (2011) The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology*, **22**, 450–456.

Jahanshad, N. (2017) Do candidate genes affect the brain's white matter microstructure? Large-scale evaluation of 6,165 diffusion MRI scans. *bioRxiv* 107987; https://doi.org/10.1101/107987, [Epub ahead of print].

Jansen, A.G., Mous, S.E., White, T., Posthuma, D. & Polderman, T.J. (2015) What twin studies tell us about the heritability of brain development, morphology, and function: a review. *Neuropsychol. Rev.*, **25**, 27–46.

Jasinska, K.K., Molfese, P.J., Kornilov, S.A., Mencl, W.E., Frost, S.J., Lee, M., Pugh, K.R., Grigorenko, E.L. *et al.* (2017) The BDNF Val66Met polymorphism is associated with structural neuroanatomical differences in young children. *Behav. Brain Res.*, **328**, 48–56.

Keller, M.C. & Miller, G. (2006) Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav Brain Sci*, **29**, 385–404; discussion 405-352.

Lubeiro, A., Gomez-Pilar, J., Martin, O., Palomino, A., Fernandez, M., Gonzalez-Pinto, A., Poza, J., Hornero, R. *et al.* (2017) Variation at NRG1 genotype related to modulation of small-world properties of the functional cortical network. *Eur. Arch. Psychiatry Clin. Neurosci.*, **267**, 25–32.

Luciano, M. (2017) 116 independent genetic variants influence the neuroticism personality trait in over 329,000 UK Biobank individuals. *bioRxiv* 168906; https://doi.org/10.1101/168906, [Epub ahead of print].

Mahmood, A., Bibat, G., Zhan, A.L., Izbudak, I., Farage, L., Horska, A., Mori, S. & Naidu, S. (2010) White matter impairment in Rett syndrome: diffusion tensor imaging study with clinical correlations. *AJNR Am. J. Neuroradiol.*, **31**, 295–299.

Mallas, E., Carletti, F., Chaddock, C.A., Shergill, S., Woolley, J., Picchioni, M.M., McDonald, C., Toulopoulou, T. *et al.* (2017) The impact of CACNA1C gene, and its epistasis with ZNF804A, on white matter microstructure in health, schizophrenia and bipolar disorder1. *Genes Brain Behav.*, **16**, 479–488.

Mefford, H.C. (2009) Genotype to phenotype-discovery and characterization of novel genomic disorders in a "genotype-first" era. *Genetics Med*, **11**, 836–842.

Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S. *et al.* (2016) Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.*, **19**, 1523–1536.

Mitchell, K.J. (2015). The genetic architecture of neurodevelopmental disorders. In Mitchell, K.J. (Ed), *The Genetics of Neurodevelopmental Disorders*. Wiley, Canada.

Moreno-De-Luca, A., Myers, S.M., Challman, T.D., Moreno-De-Luca, D., Evans, D.W. & Ledbetter, D.H. (2013) Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence. *Lancet Neurol*, **12**, 406–414.

Polderman, T.J., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M. & Posthuma, D. (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.*, **47**, 702–709.

Psaty, B.M., O'Donnell, C.J., Gudnason, V., Lunetta, K.L., Folsom, A.R., Rotter, J.I., Uitterlinden, A.G., Harris, T.B. *et al.* (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.*, **2**, 73–80.

Schmitt, J.E., Vandekar, S., Yi, J., Calkins, M.E., Ruparel, K., Roalf, D.R., Whinna, D., Souders, M.C. *et al.* (2015) Aberrant cortical morphometry in the 22q11.2 deletion syndrome. *Biol. Psychiat.*, **78**, 135–143.

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Buchel, C., Conrod, P.J., Dalley, J.W. *et al.* (2010) The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol. Psychiatry*, **15**, 1128–1139.

Shen, L., Kim, S., Risacher, S.L., Nho, K., Swaminathan, S., West, J.D., Foroud, T., Pankratz, N. *et al.* (2010) Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. *NeuroImage*, **53**, 1051–1063.

Silver, M., Montana, G. & Nichols, T.E. (2011) False positives in neuroimaging genetics using voxel-based morphometry data. *NeuroImage*, **54**, 992–1000.

Sprooten, E., Rasgon, A., Goodman, M., Carlin, A., Leibu, E., Lee, W.H. & Frangou, S. (2017) Addressing reverse inference in psychiatric

neuroimaging: meta-analyses of task-related brain activation in common mental disorders. *Hum. Brain Mapp.*, **38**, 1846–1864.
Stein, J.L., Medland, S.E., Vasquez, A.A., Hibar, D.P., Senstad, R.E., Winkler, A.M., Toro, R., Appel, K. *et al.* (2012) Identification of common variants associated with human hippocampal and intracranial volumes. *Nat. Genet.*, **44**, 552–561.
Stessman, H.A., Bernier, R. & Eichler, E.E. (2014) A genotype-first approach to defining the subtypes of a complex disease. *Cell*, **156**, 872–877.
Swartz, J.R., Waller, R., Bogdan, R., Knodt, A.R., Sabhlok, A., Hyde, L.W. & Hariri, A.R. (2017) A common polymorphism in a Williams syndrome gene predicts amygdala reactivity and extraversion in healthy adults. *Biol. Psychiat.*, **81**, 203–210.
Thompson, P.M., Andreassen, O.A., Arias-Vasquez, A., Bearden, C.E., Boedhoe, P.S., Brouwer, R.M., Buckner, R.L., Buitelaar, J.K. *et al.* (2017) ENIGMA and the individual: predicting factors that affect the brain in 35 countries worldwide. *NeuroImage*, **145**, 389–408.
Toro, R., Poline, J.B., Huguet, G., Loth, E., Frouin, V., Banaschewski, T., Barker, G.J., Bokde, A. *et al.* (2015) Genomic architecture of human neuroanatomical diversity. *Mol. Psychiatry*, **20**, 1011–1016.
Ulfarsson, M.O., Walters, G.B., Gustafsson, O., Steinberg, S., Silva, A., Doyle, O.M., Brammer, M., Gudbjartsson, D.F. *et al.* (2017) 15q11.2 CNV affects cognitive, structural and functional correlates of dyslexia and dyscalculia. *Transl Psychiatry*, **7**, e1109.
Vinkhuyzen, A.A., Wray, N.R., Yang, J., Goddard, M.E. & Visscher, P.M. (2013) Estimation and partition of heritability in human populations using whole-genome analysis methods. *Annu. Rev. Genet.*, **47**, 75–95.
Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A. & Yang, J. (2017) 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.*, **101**, 5–22.