The current state of play on the molecular genetics of depression

S. Cohen-Woods*, I. W. Craig and P. McGuffin

MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King’s College London, UK

Background. It has been well established that both genes and non-shared environment contribute substantially to the underlying aetiology of major depressive disorder (MDD). A comprehensive overview of genetic research in MDD is presented.

Method. Papers were retrieved from PubMed up to December 2011, using many keywords including: depression, major depressive disorder, genetics, rare variants, gene–environment, whole genome, epigenetics, and specific candidate genes and variants. These were combined in a variety of permutations.

Results. Linkage studies have yielded some promising chromosomal regions in MDD. However, there is a continued lack of consistency in association studies, in both candidate gene and genome-wide association studies (GWAS). Numerous factors may account for variable results including the use of different diagnostic approaches, small samples in early studies, population stratification, epigenetic phenomena, copy number variation (CNV), rare variation, and phenotypic and allelic heterogeneity. The conflicting results are also probably, in part, a consequence of environmental factors not being considered or controlled for.

Conclusions. Each research group has to identify what issues their sample may best address. We suggest that, where possible, more emphasis should be placed on the environment in molecular behavioural genetics to identify individuals at environmental high risk in addition to genetic high risk. Sequencing should be used to identify rare and alternative variation that may act as a risk factor, and a systems biology approach including gene–gene interactions and pathway analyses would be advantageous. GWAS may require even larger samples with reliably defined (sub)phenotypes.

Received 11 May 2011; Revised 12 May 2012; Accepted 16 March 2012; First published online 12 June 2012

Key words: Depression, epigenetics, gene–environment, genetics, GWAS.

Introduction

The prevalence of mental disorders in the general population is remarkably high (Kessler et al. 1994). Data collected by the World Health Organization (WHO) suggest that, as a leading cause of disability in adults, major depressive disorder (MDD) closely follows cardiovascular disease (Murray & Lopez, 1997) and is predicted to top the list by 2030. It is unfortunate, therefore, that the aetiology of affective disorders is not yet well understood, and risk genes for MDD not yet convincingly identified. This review describes heritability, the current state of play of molecular genetics in MDD, and how as a community we may go forward with our research aiming to improve molecular knowledge of the disorder.

Are genetics important in depression?

Twin studies indicate a substantial contribution of genetic and unique environmental factors to the variance observed in depression, with little or no shared environmental contribution. A meta-analysis of five twin studies, four of which were population based, suggests that genetic factors explain 37% of the variance, with unique environment accounting for 63%, and no shared environmental effects (Sullivan et al. 2000). However, heritability estimates are higher, at around 70%, in studies that either take into account diagnostic unreliability, because measurement error inflates the unique environment estimation (Kendler et al. 1993), or focus on clinically ascertained twins (McGuffin et al. 1996). These heritability estimates provide good reasoning for the attempt to identify genes that may influence susceptibility to depression, with non-shared environmental effects remaining substantial and important.

Although it has been suggested that earlier age of onset presents increased familial loading (Puig-Antich
et al. 1989; Weissman et al. 1992), this is not supported by a more recent meta-analysis by Sullivan et al. (2000) or by a later clinical twin study (McGuffin et al. 2003). Moreover, some studies suggest that pre-pubertal onset depression is largely dependent on environmental factors whereas post-pubertal onset is more genetically influenced (Weissman et al. 1987; Harrington et al. 1997). This is supported by twin (Thapar & McGuffin, 1994) and adoption studies (Eley et al. 1998).

**Finding genes involved in depression**

Linkage and association studies are complementary methods of locating susceptibility genes for MDD. Linkage can be detected over comparatively large distances but power is problematic when searching in genome-wide scans for quantitative trait loci (QTLs) with small effect sizes. By contrast, traditional association studies can detect small effects but only over very short distances. More recently, genome-wide association studies (GWAS) have become feasible, using the huge numbers of markers now available. This method combines the advantage of the breadth of linkage with the power of association, although this may be compromised by allelic heterogeneity, as discussed later.

**Linkage studies in MDD**

Despite inconsistent linkage results in MDD in earlier, mainly smaller studies (e.g. Nurnberger et al. 2001; Zubenko et al. 2002; Abkevich et al. 2003), some consistencies have started to emerge in studies using several hundreds of affected sib pairs (ASPs) (Holmans et al. 2004; McGuffin et al. 2005; Holmans et al. 2007). In the first wave of the Genetics of Recurrent Early-Onset Major Depression (GenRED) study, consisting of 415 ASPs, significant genome-wide linkage to chromosome 15q with no sex effects was reported, together with suggestive evidence for sex-specific linkage on chromosomes 6q, 8p and 17p in males (Holmans et al. 2004). It is worth noting, however, that the latter sex-specific findings may be spurious given the small male–male sample size. Two independent large-scale studies, the European–US Depression Network (DeNt) study (McGuffin et al. 2005) and a study from Utah (Camp et al. 2005), provided support for linkage in the 15q region although this was male specific in the Utah study. Similar phenotypic definitions of MDD were applied in these three studies, all investigating recurrent depression, with the GenRED and Utah studies specifically investigating early onset. The results of the final stage of a whole-genome linkage scan of 656 families in the GenRED study included fine mapping of the 15q linkage region (Holmans et al. 2007; Levinson et al. 2007; Verma et al. 2008). The findings are promising, providing what may be the first example of a genome scanning approach giving insight into the aetiology of MDD (McGuffin et al. 2007).

A second region has recently been identified that presents evidence for linkage in the combined first and second waves of the DeNt study and an independent MDD sample in the 3p24.9-26 region (Breen et al. 2011; Pergadia et al. 2011). The samples differed slightly in that DeNt was a recurrent depressive family sample with analyses restricted on impairment severity (Breen et al. 2011), and the second a depression family sample originally ascertained to explore the genetics of smoking (Pergadia et al. 2011). Although case-control association analyses in this region have not so far helped in identifying specific genes (Breen et al. 2011), the region remains of interest and the next stage for both 15q and 3p linkage regions should involve a meta-analysis, or ideally mega (combined) analyses of raw data from all available sources. It is also possible that variants contributing to the linkage peaks may be rare variants that are not detected individually by whole-genome analysis methods, and so sequencing of the candidate regions should also be carried out.

**Genetic association studies**

We are currently in an era of GWAS. However, these have only been feasible in the past few years and the majority of association studies in MDD have focused on candidate genes. In trying to establish where we need to go in the future, it is important to be aware of what has been found in the past. GWAS do not make candidate gene studies redundant because a vast majority of these studies were based on only a few single nucleotide polymorphisms (SNPs) within each candidate gene, thus not giving full coverage.

**Monoamines**

Based on what is assumed to be the mode of action of tricyclic antidepressants (TCAs) and selective serotonin reuptake inhibitors (SSRIs), obvious candidate genes for association studies in depression are those relevant to monoamine synthesis and transport together with monoamine receptors and associated G proteins.

**Monoamine oxidase**

Monoamine oxidase type A (MAOA) is a mitochondrial enzyme that has an important role in the
degradation of biological amines including 5-hydroxytryptamine (5-HT), noradrenaline and dopamine (Syagailo et al. 2001; Youdim et al. 2006). The MAOA gene is located on chromosome Xp11.23-p11.4 (Ozelius et al. 1988; Levy et al. 1989) and several polymorphisms have been identified. Two functional markers have been studied in relationship to depression; the T941G polymorphism (rs6323) does not seem to be associated with MDD (Sasaki et al. 1998; Tadic et al. 2003; Zhang et al. 2010). Many studies indicate no association with a second functional polymorphism, a variable number tandem repeat (VNTR), and affective disorder or specifically depression (e.g. Kunugi et al. 1999; Christiansen et al. 2007). However, positive reports include allele and homozygous genotype association of the high-activity variants with MDD, particularly in females (Schulze et al. 2000; Yu et al. 2003; Rivera et al. 2009), and association with subtypes of MDD (Gutierrez et al. 2004). There has also been a conflicting report associating the low-activity allele with depressive symptoms (Brummett et al. 2007).

**Serotonin**

**The serotonin transporter**

The serotonin transporter (5HTT) gene located at 17q11.1-q12 has been studied extensively, with two well-characterized polymorphisms the focus of much depression-related genetic research. At the 5′ flanking regulatory region of the human 5HTT gene there is a functional 44-base-pair (bp) insertion/deletion polymorphism, known as the serotonin-transporter-linked polymorphic region, 5HTTLPR (Heils et al. 1996), conferring long and short alleles associated with differential 5HTT expression and 5-HT reuptake in lymphoblastoid cell lines (Lesch et al. 1996).

Despite its strong candidacy, the numerous association studies investigating the 5HTTLPR and depression have reported equivocal results (e.g. Collier et al. 1996; Furlong et al. 1998b; Minov et al. 2001; Hauser et al. 2003). This could be a consequence of small sample sizes and several meta-analyses have been conducted in an attempt to circumvent this problem. The first included a total of 275 depressed patients and 739 controls of Caucasian origin, and demonstrated an association with the short allele and depression (Furlong et al. 1998b). This has since been supported by two further meta-analyses (Lotrich & Pollock, 2004; López-León et al. 2008); however, two other meta-analyses present no evidence for association (Anguelova et al. 2003; Willis-Owen et al. 2005). This may be explained by the heterogeneity of inclusion criteria for the analyses.

Although the 5HTTLPR has been reported to affect transcriptional activity alone, other studies indicate a more complex situation (e.g. Nakamura et al. 2000). An SNP, rs25531 (A/G), has been described that influences transcriptional activity as part of a haplotype with the 5HTTLPR (Nakamura et al. 2000; Wendland et al. 2006). This G substitution in combination with the long allele (L_C) seems to reduce the expression to the same level as the short allele, thereby creating two functional haplotype groups: L_A versus L_C, S_A and S_C. Relatively little research incorporating this SNP has been published to date. One study reports a female-specific association with melancholic depression specific to L_A as the risk haplotype (Baune et al. 2008) whereas another study reports a protective effect of this haplotype on treatment-resistant depression (Bonvicini et al. 2010). The situation may be further complicated by the apparent functional impact of additional proximate SNPs (Martin et al. 2007).

A 16/17-bp VNTR (Cowen & Charig, 1987; Kaiser et al. 2001) in the second intron of the 5HTT gene (STin2) has also been investigated in association with MDD. Again, association studies have been equivocal; the most promising association is with the 9-repeat (e.g. Battersby et al. 1996; Ogilvie et al. 1996; Bozina et al. 2006), although several studies suggest this is not the case (e.g. Kunugi et al. 1997; Furlong et al. 1998b). Three meta-analyses suggest that no significant association exists between the VNTR alleles and/or genotypes and depression (Furlong et al. 1998b; Anguelova et al. 2003; López-León et al. 2008), but samples analysed to date have had low numbers.

**The serotonin receptors**

The serotonin receptor 1A is encoded by an intronless gene (HTR1A) located at 5q11.2-q13 (Melmer et al. 1991). Most studies indicate little evidence for association of this locus as confirmed by a meta-analysis of the marker SNP rs6295 (López-León et al. 2008). The serotonin receptor 1B gene (HTR1B) maps to human chromosome 6q13, a region for which sex-specific (male) linkage to depression and/or anxiety has been reported (Holmans et al. 2004; Nash et al. 2004). The variant G861C (rs6296) is a functional polymorphism (Maura et al. 1993); however, no significant association has been found directly with MDD although it has been suggested that rs6296 may only be associated with a more severe (i.e. recurrent) form of depression (Huang et al. 2003).

Studies investigating the serotonin receptor 2A gene (HTR2A) at 13q14-q21 have also been inconclusive. Two meta-analyses indicate no association with depression (Anguelova et al. 2003; López-León et al. 2008). The serotonin receptor 2C gene (HTR2C) is
situated on chromosome Xq24, with a functional polymorphism (rs6318, a Cys23Ser substitution) in the hydrophobic region of the receptor (Lappalainen et al. 1995; Quested et al. 1999). Although no convincing evidence for association exists (Lerer et al. 2001; Koks et al. 2006), given the sex imbalance observed in depression, further research investigating this X-linked gene is merited.

Other serotonin receptor genes studied include HTR3A and HTR3B, which are tandemly positioned at 11q23 (Miyake et al. 1995; Davies et al. 1999). Little evidence for association is observed, although three-marker haplotypes spanning a seven-marker region in the HTR3B gene are reported to be associated in the Japanese female population (Yamada et al. 2006) (see online Supplementary Table 1). Other receptor genes have also been examined, including HTR5A (7q36.1) and HTR6 (1p35-36), but no strong evidence for association has emerged.

**Tryptophan hydroxylase (TPH)**

TPH is rate limiting in the biosynthesis of 5-HT (Priestley & Cuello, 1982). The TPH1 gene has been mapped to human chromosome 11p15.3-p14 (Craig et al. 1991). An association with the C/C genotype of the A218C (rs1800532) SNP and depression has been reported (Tan et al. 2003), particularly with severe depression (Viikki et al. 2010); however, an increase in the frequency of the A allele in depressed patients has also been observed (Tsai et al. 1999; Wang et al. 2011). Population stratification could account for this as the former two studies used Singaporean and Finnish subjects, and the latter Chinese and Taiwanese. However, multiple negative studies have also been published (e.g. Frisch et al. 1999; Gizatullin et al. 2006). A meta-analysis including both the positive and negative studies suggests that the A128C SNP is not associated with MDD (López-León et al. 2008). A second SNP in intron 7, C779A (rs1799913), has been associated with depression (Gizatullin et al. 2006), although this has not been replicated (Koks et al. 2006).

A second TPH isoform was first identified in the mouse (Tph2) and shares 72% sequence homology with the TPH1 gene (Walther & Bader, 2003). In humans, TPH2 at 12q21 falls within a previously reported MDD linkage region (Abkevich et al. 2003). There was considerable excitement when a loss-of-function mutation, R441H, was described and reported to be highly associated with MDD (Zhang et al. 2005); however, this variant seems to be extremely rare and has not been observed in multiple independent studies since the initial report (e.g. Glatt et al. 2005).

**Noradrenaline**

**The noradrenaline transporter**

Serotonin–norepinephrine reuptake inhibitor (SNRI) antidepressant medication is known to act on the noradrenaline transporter (SLC6A2) and a reduction in its density in the locus coeruleus has been reported in depressed patients (Klimk et al. 1997). The SLC6A2 gene maps to chromosome 16q12.2 (Porzgen et al. 1995). Many studies investigating different populations have failed to detect an association using a promoter polymorphism, T-182C (rs2242446), the synonymous variant G1287A (rs5569) and/or the non-synonymous exonic variant C296T/Thr99Ile and depression (e.g. Owen et al. 1999; Inoue et al. 2007), whereas two studies on –182T/C report an association but with different genotypes (Inoue et al. 2004; Ryu et al. 2004). Unsurprisingly, a meta-analysis of the –182T/C locus presents no evidence for association with MDD (López-León et al. 2008). However, a study that sequenced the exons to identify all SNP variations does report evidence for association in a Mexican-American population, highlighting the importance of a comprehensive approach (Dong et al. 2009).

**The adrenaline and noradrenaline receptors**

The α2A- and α2C-adrenergic receptors, unlike α2B, are both expressed in brain and have been implicated in stress (e.g. Fuchs & Flugge, 2003) and MDD (e.g. Ordway et al. 2003). Nevertheless, no association is reported with either the ADRA2A –1291C/G polymorphism or the Gly389Arg ADRB1 polymorphism, and MDD or mood disorder (Ohara et al. 1998; Zill et al. 2003). A comprehensive survey of 24 markers in nine adrenergic receptor genes, including three that fall within linkage regions for recurrent MDD [ADRB1 (5q23-q33.3), ADRB3 (8p12-p11.2) and ADRB1 (10q24-q26)] (Zubenko et al. 2003), undertaken in a Hungarian cohort of early-onset depression (before the age of 15 years) failed to find strong evidence for association (Burcescu et al. 2006).

**Dopamine**

**The dopamine receptors and transporter**

Only two published studies have investigated the dopamine D1 receptor (DRD1) gene, located at 5q35.1, in association with MDD. One reported on 10 SNPs (Koks et al. 2006) and the other on four (Garriock et al. 2006); three were investigated by both groups (~1251G/C, ~800T/C, ~486G/A). Neither study finds significant association. Similarly, no positive association has been reported for the functional Ser311Cys
substitution (rs1801028) in the DRD2 gene (11q22.3-q23) and MDD (e.g. Manki et al. 1996; Koks et al. 2006), or the −141C insertion/deletion in a relatively small study (n = 128 patients and 262 controls) (Furlong et al. 1998a). Investigation of nine additional SNPs in the DRD2 gene further indicate no role of this gene in MDD susceptibility (Koks et al. 2006). A Ball polymorphism and a Ser9Gly substitution (rs6280) in the first exon in DRD3 have been investigated; one study presented significant association with the Gly9 allele and homozygote genotype (Dikeos et al. 1999) but this has not been replicated in larger studies (e.g. Manki et al. 1996; Garriock et al. 2006). A meta-analysis also suggests that there is no association (López-León et al. 2008). A 48-bp VNTR has been described in exon 3 of the DRD4 gene (11p15.5), which has been significantly associated with depression (Manki et al. 1996), although again this has not been replicated in other studies (e.g. Oruc et al. 1997; Frisch et al. 1999).

The dopamine transporter gene (DAT1 or SLC6A3), located on 5p15.33, has also been investigated in MDD. Although studies investigating a 40-bp VNTR in the three prime untranslated region (3′ UTR) report no evidence for association (Manki et al. 1996; Frisch et al. 1999), a meta-analysis comparing the 9/10 genotype with the 10/10 genotype suggests the former to be associated with MDD (López-León et al. 2008).

Beyond the monoamine hypothesis

The hypothalamic–pituitary–adrenal (HPA) axis and depression

Given the close relationship between stress and depression (Charney & Manji, 2004), genes associated with the HPA axis are attractive candidates. MDD has been associated with HPA axis hyperactivity (Plotsky et al. 1998) and research demonstrates that HPA axis dysfunction is genetic and increases the risk of depression (e.g. Modell et al. 1998). Genetic research to date investigating candidate genes in this system are summarized in the online Supplementary Table 2. Although it is clear that no consistent association is observed, studies are few and limited variously by heterogeneous genetic variation, small sample sizes and/or imprecise phenotypes.

Brain-derived neurotrophic factor (BDNF)

Chronic stress has also been associated with hippocampal neuronal atrophy in animal studies (Sapolsky et al. 1990) and found to reduce BDNF expression in hippocampal neurons (Smith et al. 1995). Antidepressants and electroconvulsive therapy (ECT) have been found to inhibit this stress-induced response (e.g. Duman, 1999). The BDNF gene is mapped at 11p13/11p14 (Hanson et al. 1992). Depression research has concentrated on one functional polymorphism (rs6265), sometimes coupled with a GTn microsatellite polymorphism and a range of other SNPs (see online Supplementary Table 3). Although reports have been conflicting, the largest case-control study published to date, while failing to find convincing single-marker association, reports a haplotype association with three markers (rs6265-GTn, repeat-rs988748) that is also observed in a second sample (Schumacher et al. 2005). This has not been replicated in a larger case-control sample (n = 1450 cases and 850 controls; Cohen-Woods, unpublished observations). The involvement of the BDNF gene in depression remains controversial, with meta-analyses mirroring the landscape of the inconsistent individual studies (see online Supplementary Table 3), presenting different findings depending on the studies included and whether gender and/or ethnicity is considered (Gratacos et al. 2007; Verhagen et al. 2010).

GWAS

GWAS using micro-arrays enable the order of one million SNPs to be genotyped simultaneously for each individual. These studies have the potential to identify candidates in underexplored pathways relating to depression. It remains to be seen if the success achieved with other common traits and disorders such as type 2 diabetes (Zeggini et al. 2007) and rheumatoid arthritis (Thomson et al. 2007), where GWAS have led to the discovery of novel susceptibility genes, can be replicated. In the first report of this kind, the candidacy of a gene implicated in monoaminergic transmission (piccolo, PCLO) emerged after re-examination of the original cohorts concentrating on samples of greatest similarity (Sullivan et al. 2009). A p value of 6.4 × 10^−8 was achieved for a non-synonymous SNP near a calcium-binding domain. However, although replication was lacking within the study, a more recent population-based study does replicate an association with the PCLO gene and depressive disorders, although it was not a whole-genome association study (Hek et al. 2010). Another genome-wide study implicates a novel locus, BICC1, but this failed to be confirmed in independent replication samples (Lewis et al. 2010). Further whole-genome studies investigating MDD fail to identify any loci or replicable loci (Muglia et al. 2010; Rietschel et al. 2010; Terracciano et al. 2010; Shi et al. 2011; Wray et al. 2012). A recent meta-analysis has included data from three of these studies (Sullivan et al. 2009; Muglia et al. 2010; Shi et al. 2011) and reported three intronic SNPs in ATP6V1B2, SP4 and GRM7 as most associated, but
with p values just short of genome-wide significance (Shyn et al. 2011). In addition to meta-analysis it is possible to perform a ‘mega-analysis’ that pools samples together: this is the approach undertaken by the Psychiatric GWAS Consortium (PGC). In their discovery sample 9240 affected MDD individuals and 9519 controls were included, with independent replication samples totalling 6783 affected MDD individuals and 50 695 controls. In total, more than 1.2 million SNPs were analysed, with none reaching significance in the MDD groups (Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, 2012).

Why has there been so little replication?

Despite significant heritability, failure of replication is a persisting feature and can be attributed to a variety of plausible explanations including sample size, heterogeneity, publication bias, multiple-testing issues, and insufficient quality control. These are discussed briefly in the following sections.

Size of our samples

MDD is a complex multifactorial disease, where it is probable that variants of multiple genes of small effect are being sought. This means that samples must be very large to detect any association, a criterion that has rarely been achieved and that frequently results in false positives, excluding the possibility of replication, thereby creating an unfortunate non-self-fulfilling prophecy. With the advent of whole-genome studies, the issue of power and sample size becomes ever more significant, as larger numbers are required to provide sufficient power to detect association because of the impact of multiple testing in addition to the small effect sizes of risk loci. GWAS had greater numbers than those studies often used in earlier candidate gene studies, but still the numbers do not seem to be sufficient to identify consistently associated loci. The PGC has tried to address this issue with the mega-analysis described earlier, but even with such large numbers little of significance has been identified. Nonetheless, there is a strong argument that we must not give up on GWAS because, although numbers are greater than previously studied, much greater numbers are probably required to detect common variation of small effect in psychiatric disease (Sullivan, 2012).

Population stratification

The frequencies of many markers are sensitive to population differences and ethnic stratification of samples can confound data considerably. Even a small admixture can confound results, particularly in haplotypic analyses when considering multiple markers in combination, and although some studies check for stratification, this is not always reported. It has been reported that homogeneous control Caucasian populations with comparable allelic frequencies can demonstrate different underlying linkage disequilibrium (LD) and haplotype structures within Europe (Mueller et al. 2005). This may create difficulty in the replication of specific haplotype associations or even single-marker associations, where the true associated locus is in fact in LD with the single marker and so not detected across populations because of differing population structures.

Multiple testing

A consequence of genome-wide capabilities is the necessity to run hundreds of thousands of statistical tests simultaneously. Although genome-wide significance thresholds have been established in the literature, there is the real risk of false negatives (Panagiotou et al. 2012). Thus, sufficiently powered candidate gene studies still have their role within the literature, but authors should attempt to locate multiple samples within which to replicate findings. Similarly, targeted replication can be used to support evidence for a genome-wide associated locus from one study in independent samples. This reduces the multiple testing burden in the replication sample(s). A method that attempts to circumvent, or at least reduce, the problem with multiple testing and assuming independence is by taking a systems biology approach in pathway analysis. To our knowledge there is only one study to date applying this approach in the literature. Although producing interesting results, with pathways including neurotransmitter, immune and inflammatory responses identified, there were in total 17 significantly enriched pathways in MDD; rather too many for specific studies (Kao et al. 2012). Therefore, the key findings require replication. The PGC is also working on a huge collaborative pathway analysis effort, despite the lack of individual studies published.

Heterogeneity

Problems with replication may also be a consequence of phenotypic and genetic heterogeneity. Whereas the former is self-explanatory, genetic heterogeneity can take two forms: locus and allelic heterogeneity. Locus
heterogeneity occurs when different genes (or combinations of genes) influence disease risk between individuals. By contrast, allelic heterogeneity refers to different variants within the same gene influencing disease risk. Allelic heterogeneity has often been cited as a reason for the lack of specific SNP replication across studies in which different alleles of one gene track the disorder. The existence of multiple rare alleles of high impact is a plausible explanation in such cases. Methods to address phenotypic heterogeneity can include using sub-(or endo)phenotypes, such as specific types of depression (e.g. post-natal, melancholic). Deep sequencing enables us to address allelic heterogeneity, targeting genetic regions implicated by either linkage or association studies that may harbour rare or common variations predisposing to MDD. Sequencing would capture variation not genotyped by these studies but linked to the associated locus/region.

**Type of variation studied**

The discrepancy between results may, however, go beyond basic statistical and methodological issues. The type of variation being studied, genetic and non-genetic, also has the potential to explain in part the problems with replication. This includes environmental factors, epigenetics, copy number variation (CNV) and rare variants, all discussed in brief here.

Further to establishing strong genetic influence in depression, quantitative genetic studies have found substantial influence of an individual’s unique environment (Kendler, 1996). Consideration of environmental factors is often ignored in psychiatric genetics and, in such circumstances, it might be unsurprising that consistent results are not achieved. In addition to simply co-acting in an additive fashion, genetic and environmental factors affecting the aetiology of MDD may covary and interact with one another. It has been postulated that one-third of the variance observed between stressful life events and depression is attributable to gene–environment (G–E) correlation (Kendler et al. 1999). Evidence of G–E interaction in depression was first demonstrated in a study by Caspi et al. (2003), in which the 5HTTLPR s allele only conferred susceptibility to depression if the carrier had been exposed to stressful life events. This has since been replicated (e.g. Eley et al. 2004; Willhelm et al. 2006), but not consistently (e.g. Fisher et al. 2012). Two meta-analyses failed to support evidence for the G–E interaction, including nine and 14 of the 54 G–E 5HTTLPR studies published to date respectively (Munafo et al. 2009; Risch et al. 2009). By contrast, the most recent meta-analysis, which included all but one study, shows that, depending on the type of variation studied (i.e. if limited to childhood maltreatment or physical illness), there is strong evidence for G–E interaction. Consideration of adult life events, or brief life events, did not support G–E interaction, showing the importance of the variables considered (Karg et al. 2011). The type of measurement (i.e. subjective versus objective) has also been reported to have a potentially significant impact (Uher, 2011). This is exemplified by our own depression study, which found no evidence for G–E interaction when investigating adult life events (Fisher et al. 2012) but strong evidence with childhood maltreatment factors (Fisher et al., in press). However, a recent study aimed directly at replicating the original 5HTTLPR interaction with life stress reported by Caspi et al. (2003) failed to find any evidence for interaction using a variety of stress and outcome measures, including childhood maltreatment (Fergusson et al. 2012). Although this paper has not been included in any meta-analysis to date, there is an ongoing international effort to perform coordinated reanalysis of primary data from this and many other published and unpublished datasets to address the complex question of heterogeneous environmental and outcome measures in G–E 5HTTLPR studies. This highlights that, although there is some optimism in this field of research, some caution is also necessary (Brown, 2012). Nonetheless, as many candidate gene effects may be influenced by stress, the inclusion of measures of stress-related variables and other environmental factors in research designs is recommended and may soon become the norm; however, the type of variable collected is important in consideration of research designs.

Epigenetic factors have also not been widely investigated in depression, and it has been suggested that a traditional DNA sequence-based approach may not fit many common aetiological characteristics seen in complex disorders (Petronis, 2001). These include the relatively high incidence of discordance among monozygotic twins, the comparatively delayed onset, the frequently reported sex effects, and parent-of-origin effects (Petronis, 2001). Further to this, epigenetics provides a mechanism by which the environment may interact with an individual’s genetic vulnerability to precipitate MDD, although epigenetics may be an important factor independent of environmental influence. It seems that no study has yet investigated whether depression exhibits parent-of-origin effects, which, if they exist, would increase the rationale for epigenetic investigation. There is some debate regarding the utility of epigenetic study when we are unable to access the tissue of interest (i.e. brain) in living patients. However, establishing whether epigenetic biomarkers exist in peripheral tissues is important; herein lies the potential to identify
individuals at risk of developing MDD and/or identifying new molecular pathways for improved pharmacotherapeutics.

A further avenue that looks promising is research into CNVs in the human genome. In schizophrenia there seems to be an increased burden of deletions (International Schizophrenia Consortium, 2008), which is also observed in autism (Walsh et al. 2008) and in bipolar disorder (Zhang et al. 2009). In bipolar disorder, however, the findings have been less consistent, with other studies failing to report such evidence for association with CNVs (McQuillin et al. 2011). In depression, only one study has investigated CNVs on a genome-wide level and reported evidence for an increase in burden in the MDD cases (Rucker et al. 2011). Of particular interest, two control groups were investigated: one psychiatrically screened control group and the other sourced from the general population (the Wellcome Trust Case Control Consortium 2, WTCCC2). The CNV burden was found to be intermediate between the MDD and the psychiatrically screened control group in the general population controls, indicating that less deletions could in fact be an indicator for well-being; however, this requires independent replication. Another genome-wide CNV study supported the implication of CNVs in MDD but did not analyse the data in the same manner as a majority of studies to date that have focused on burden rather then specific CNV association (Glessner et al. 2010).

Finally, in all of the SNP studies and the recent GWAS being published, microsatellites have been generally neglected. These markers are valuable in genetic analysis, having a wide range of population ancestries and in some cases a greater reach of LD. Their impact on the genome may present important structural and functional implications beyond those observed for SNPs. Furthermore, studies have focused on common variation, following the common-disorder common-variant hypothesis; it is possible that the levels of variation implicated in MDD might be rarer than originally anticipated and so have been missed by our candidate and GWAS methods. Sequencing has the potential to address both issues of detecting non-SNP variation, and more rare variants.

Conclusions

In summary, large sample sizes (between 2000 and 10,000) and precise phenotypic definitions may be necessary to detect variants with small effect sizes. This issue is being addressed by the work of the PGC, and the results in MDD will soon be published. Data already collected from GWAS can be applied to other approaches that apply more of a systems biology approach, such as gene–gene interactions and pathway analyses. Regions that have been identified either by linkage studies or by GWAS would benefit from targeted deep sequencing to determine whether there is a rare variation that might contribute to MDD risk. The variation identified should then be genotyped again in independent samples that may have a different underlying LD structure, and so may not have initially presented evidence on the basis of the GWAS data. Subphenotyping to reduce heterogeneity within samples also has the potential to yield loci for more specific aspects, or subgroups, of MDD. Environmental measures should be included in future research designs. Greater emphasis should be placed on the precise nature of environmental measures in molecular behavioural genetics, where possible (including their potential epigenetic consequences), to allow the identification of individuals at high environmental risk in addition to high genetic risk. We accept that not every study is able to address every issue, and each research group should identify what it is their samples might best address; some samples are well disposed to gene–environmental study whereas others are better suited to subphenotyping, epigenetic study and/or sequencing. A combination of all these considerations should lead to more holistic, reliable and perhaps replicable data, reflecting the genetic underpinning of depression and informing the design of future sample ascertainment.

Supplementary material

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0033291712001286.

Acknowledgements

We are grateful to E. Halton for excellent assistance with the preparation of this manuscript.

References


Porzgen P, Bonisch H, Bruss M (1995). Molecular cloning and organization of the coding region of the human...


Sullivan PF (2012). Don’t give up on GWAS. *Molecular Psychiatry* 17, 2–3.


Sullivan PF, Neale MC, Kendler KS (2000). Genetic epidemiology of major depression: review and


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.