

Aetiological stratification as a conceptual framework for gene-by-environment interaction research in psychiatry

Ruud van Winkel^{1,2*}

¹ Department of Psychiatry and Psychology, School for Mental Health and Neuroscience, EURON, Maastricht University Medical Centre, PO Box 616 (Vijv1), 6200 MD Maastricht, The Netherlands

² University Psychiatric Center Katholieke Universiteit Leuven, campus Kortenberg, Leuvensesteenweg, Kortenberg, Belgium

It has been argued that gene-by-environment interactions ($G \times E$) research is unlikely to progress knowledge about psychiatric disorders, in contrast to genome-wide association (GWA) studies. However, $G \times E$ approaches are not alternatives for gene-hunting but a way to identify genetic and biological mechanisms within subgroups of patients exposed to a similar aetiological (environmental) factor via a process of 'aetiological stratification'. This is important as diagnostic categories targeted by GWA studies are inherently heterogeneous and lack biological validity. Aetiological stratification builds on examining possible phenotypic and/or molecular specificity associated with exposure to the environmental factor across multiple potentially relevant disorders, combined with efforts to identify an underlying biological substrate. $G \times E$ hypotheses within this framework investigate (1) which genes influence the degree to which individuals develop identified biological alterations that link environmental exposure to specific phenotypic and/or molecular characteristics within or across psychiatric disorders and (2) which genes are implicated in determining the development of psychopathology once this biological alteration has been brought about. As gene-hunting is not a goal in itself, the examination of pathway and/or polygenic risk scores may be more informative than the examination of individual markers, at the same time reducing multiple testing and the associated risk of spurious findings.

Received 10 July 2014; Accepted 1 August 2014; First published online 10 September 2014

Key words: outcome studies, risk factor, schizophrenia, schizoaffective disorder.

Introduction

For many years, it has proven extremely difficult to identify genes for psychiatric disorders (Sullivan, 2008). Trading the candidate gene approach for methods that systematically interrogate the human genome, in combination with stringent control for multiple testing and replication in independent cohorts, has now resulted in the first replicable findings in the field of psychiatric genetics (Sullivan *et al.* 2012).

In contrast, the field of gene–environment interaction ($G \times E$) research has yet to produce its first unambiguously replicated finding. Although there are a number of promising findings (for review see (Caspi *et al.* 2010; Modinos *et al.* 2013; van Winkel *et al.* 2013; van Winkel & Kuepper, 2014)), recent work has also demonstrated considerable methodological issues. First and foremost, there is evidence for widespread publication bias in the

$G \times E$ literature (Duncan & Keller, 2011). Second, while at first glance several findings have been replicated, many of these replications used different outcome variables and study designs, and very few studies have tried to conduct an exact replication study. Most of the exact replication studies produced negative results (Duncan & Keller, 2011; van Winkel & Kuepper, 2014).

Some investigators have therefore argued that $G \times E$ research is unlikely to progress knowledge about psychiatric disorders (Zammit *et al.* 2010). Arguments put forward are that approaches studying main effects of genes (such as genome-wide association (GWA) studies) are not plagued by difficulties to accurately measure environmental exposures, have considerably more statistical power and will require less control for multiple testing given the considerable number of possibly relevant environmental factors in $G \times E$ research. A reconsideration of the place of $G \times E$ research in a time-frame where genome-wide approaches have become the mainstay of psychiatric genetics is thus needed. In this editorial, I will argue that $G \times E$ approaches are not alternatives for GWA studies in the hunt for genes but

*Address for correspondence: R. van Winkel, Department of Psychiatry and Psychology, School for Mental Health and Neuroscience, EURON, Maastricht University Medical Centre, PO Box 616 (Vijv1), 6200 MD Maastricht, The Netherlands.
(Email: ruud.vanwinkel@maastrichtuniversity.nl)

should be regarded as a method to identify genetic and biological mechanisms within subgroups of patients characterised by similar aetiological exposures, and that applying this line of reasoning facilitates many of the most important choices in this field of investigation.

The ‘reification’ of Diagnostic and statistical manual of mental disorders (DSM): the one disease principle, applied to schizophrenia

An important issue is that GWAS treats disorders as uniform diseases, as things that exist in real life (‘reification’) rather than a constellation of commonly co-occurring symptoms without a known biological basis. Many findings argue for substantial phenotypic and aetiological heterogeneity in psychiatric disorders, much of the evidence for aetiological heterogeneity coming exactly from GWA studies. For example, GWA studies have found considerable genetic overlap between schizophrenia and bipolar disorder, depression and weaker but significant overlap with autism spectrum disorders and attention deficit hyperactivity disorder (ADHD) (Purcell *et al.* 2009; Cross-Disorder Group of the Psychiatric Genomics Consortium *et al.* 2013). Moreover, GWA studies identified multiple copy number variants that are associated with schizophrenia, most of which were also associated with other neurodevelopmental disorders, such as mental retardation, epilepsy, autism and other disorders (Sebat *et al.* 2009).

In addition to findings from genetic studies, there are multiple environmental factors that are associated with schizophrenia, and the evidence suggests that at least some of them are likely causally associated to the disorder. Current aetiological models of psychosis therefore acknowledge the inherent aetiological heterogeneity of ‘schizophrenia’ and have postulated that the biological basis that psychotic disorders may have in common is (presynaptic) dopamine dysregulation, giving rise to perceptual abnormalities and delusion formation, that is dopamine dysregulation is the ‘final common pathway to psychosis’ (Howes & Kapur, 2009).

The unique disease principle and its application in other branches of medicine

Aetiological heterogeneity is not specific to psychotic disorders, or psychiatric disorders in general. In fact, it is likely the norm rather than the exception in complex diseases. Indeed, human complex diseases typically result from a complex interplay between the personal genome and highly individual exposures to the environment, leading to unique alterations in the epigenome, transcriptome, proteome and metabolome (Ogino *et al.* 2011). In psychiatric disorders, these

factors tap onto inherent temperaments, leading to individual personality profiles that interact with environments, feelings and emotions to produce a set of symptoms that surely must be unique, and nominally distinct from what is considered to be the same disease process in other individuals. This concept is called the ‘unique disease principle’ (Ogino *et al.* 2013) and poses considerable challenges to traditional epidemiological research, as traditional epidemiology assumes that we can predict disease occurrence and evolution by studying groups of individuals suffering from a disorder with the same name.

In the field of oncology, where similar challenges in terms of disease heterogeneity apply, researchers have coined ‘molecular pathological epidemiology’ (MPE) as a means of dealing with heterogeneity (Ogino *et al.* 2011, 2013). In MPE, a known or suspected aetiological factor is examined in relation to specific molecular alterations (as evident in the examinations of tissues by a pathologist, hence ‘pathology’), in order to gain insight in the pathophysiology of tumour genesis (Ogino *et al.* 2011, 2013). In other words, stratification according to involved aetiological factors and studying them in relation to specific relevant alterations, in order to identify clinically meaningful subtypes. This approach tries to find a middle ground between the ‘one disease principle’ and the ‘unique disease principle’, by asserting that it is possible, to some extent, to predict the onset, persistence and progression of specific disease subtypes that were exposed to a similar aetiological factor.

The advantages of such a stratified approach are obvious: if relationships between aetiological factors and specific disease characteristics or subtypes can be uncovered, this supports the causality of the relationship as well as the assumption of aetiological heterogeneity. Moreover, this approach allows us to target specific subgroups in terms of personalised prevention, treatment and lifestyle interventions.

Aetiological stratification in psychiatry

The application of aetiological stratification in psychiatry is not new. A field of enquiry that uses aetiological stratification is, for example, represented by researchers studying 22q11.2 deletion syndrome. It is well-established that this particular deletion increases risk for schizophrenia (approximately 25% of 22q11.2 deletion carriers will develop schizophrenia). Research in this subgroup has yielded evidence for specificity: for example, evidence suggests that the auditory hallucinations that are present in this syndrome may be the result of disrupted synaptic transmission between the thalamus and the auditory cortex, mediated in part via a gene disrupted by the

deletion –Dgcr8–, whereas there is no compelling evidence that a similar mechanism is involved in other patients with schizophrenia (Chun *et al.* 2014).

Work from my group, using an *environmental stratification* approach, has provided robust and replicated evidence that exposure to childhood trauma is associated with a specific admixture of affective, anxious and psychotic symptoms (van Nierop *et al.* 2014b). This admixture was present in patients with a mood disorder (depression and bipolar disorder), in patients with an anxiety disorder, as well as in patients with schizophrenia. Moreover, the same admixture was also shown at the subclinical level in a representative general population sample (van Nierop *et al.* 2014b). We also examined functional correlates in these patient samples and reported that patients with childhood trauma-related admixture of symptoms were characterised by significantly worse functioning than non-traumatised patients, or traumatised patients without symptom admixture, in terms of working situation, relationships, living situation, symptom severity, substance abuse co-morbidity, social engagement and interaction, and quality of life. Functioning was especially poor in patients with schizophrenia with childhood-related symptom admixture (van Nierop *et al.* 2014a). Together, these data suggest that childhood trauma is an important aetiological factor in the developmental trajectory towards different psychiatric disorders, characterised at the phenotypic level by an admixture of affective, anxiety and psychotic symptoms.

Implications for G × E research

So, how is all this relevant to G × E research? The application of genome-wide methods has shown the productiveness of this approach for gene-finding efforts in all areas of medicine. Given that GWA studies do not require comprehensive assessment of phenotypic and environmental characteristics, provide more statistical power than G × E approaches and are statistically less challenging, it seems fair to assume that GWA studies are best suited for gene-finding of psychiatric disorders. It is however important to realise that GWA studies have typically found many single nucleotide polymorphism (SNPs) with small individual effects; that is, they probably identify SNPs with small effects on intermediate phenotypes that are relevant to all (or most) patients with a given disorder, or identify SNPs that have larger effects in subgroups but smaller to absent effects in other subgroups. This conclusion follows logically from the fact that GWA studies, because of their design, do not take into account aetiological heterogeneity. This is not necessarily problematic as long as GWA studies are complemented by

other approaches that deal with the inherent aetiological heterogeneity and lack of biological validity. This is where G × E designs can prove to be useful: in identifying genetic/biological mechanisms operating in subgroups of patients stratified according to the involved aetiological factors, that is by studying how genetic markers influence the effects of identified environmental agents within and/or across diagnostic categories. These interactions may be modelled along the entire developmental pathway, as well as on the course and outcome within given subgroups. Work in the field of post-traumatic stress disorder (PTSD) has shown the productiveness of such an approach. By stratifying PTSD patients, researchers have been able to show distinct genome-wide epigenetic profiles in subgroups with and without exposure to childhood trauma (Mehta *et al.* 2013), as well as site-specific methylation changes in the FKBP5 gene (Klengel *et al.* 2013). They also were able to show that differences in the vulnerability to develop these methylation changes were dependent on certain SNPs within the FKBP5 gene (Klengel *et al.* 2013). Importantly, a framework of ‘aetiological stratification’ also provides guidance on some of the most important considerations in G × E research: which exposures to select, which phenotypes to target, how to model the ‘G’ component into G × E hypotheses and how to reduce the risk for spurious findings.

Which exposures and phenotypes to target

Many environmental factors are associated with psychiatric disorders, and many more may be plausibly linked to ‘intermediary’ phenotypes that are relevant for psychiatric disorders. Not all of these exposures, however, may be ideally suited for G × E research. In order to reduce the risk of excessive multiple testing, priority should be given to environmental factors robustly associated with one or multiple psychiatric disorders, preferably confirmed by meta-analyses. Second, reported effect sizes should be large enough to allow for meaningful stratification. Third, there should ideally be sufficient proximity to an underlying biological mechanism. For example, the mechanism via which growing up in an urban environment acts upon the biology to increase psychosis risk is completely unknown and may involve pollution, social anonymity, lack of green space, stress or a combination of these and/or other factors, making it a less ideally suitable exposure for G × E studies.

With regard to the phenotypes to target, careful examination of the epidemiological patterns that are evident for different putative aetiological factors should guide the choice of the best possible approach. For example in the case of cannabis, evidence suggests that exposure is specifically associated with psychosis,

or at least considerably more strongly and convincingly than, for example, affective disorders (Moore *et al.* 2007). In this case, restricting the analyses to psychosis, or neurobiological mechanisms underlying psychosis such as pre-synaptic dopamine hyper-reactivity, would therefore probably be the most appropriate choice. In contrast, exposure to childhood trauma is associated with a range of different disorders. In this instance, I would argue that careful examination of the epidemiological patterns across and within diagnostic categories is necessary to identify the most appropriate approach to phenotype selection. This means that, in contrast to what is current practice, not diagnostic categories should be leading in framing G×E hypotheses ('does exposure×interact with gene Y to increase risk for disorder Z?'). Rather, in the framework of aetiological stratification, these questions should be preceded by:

- (1) examining the epidemiological pattern that is evident in terms of associations across multiple potentially relevant disorders ('associated with one or multiple disorders?');
- (2) looking for evidence for possible specificity in terms of phenotypic presentation or molecular markers across these categories ('is there evidence for phenotypic or molecular specificity in subgroups exposed to the aetiological factor?');
- (3) efforts to identify biological mechanisms within groups with both exposure to the environmental factor and identified specific phenotypic or molecular characteristics ('is there a biological substrate linking the environmental exposure to the specific phenotypic or molecular characteristics?').

The finding of a childhood trauma-related admixture phenotype, as described above (van Nierop *et al.* 2014b), is an example of such an approach, although further fine-mapping of the phenotypic characteristics as well as the underlying biology is necessary. In any case, once these questions have been examined, G×E hypotheses follow logically, that is (1) which genes influence the degree to which individuals develop the identified biological alterations that link exposure to the aetiological factor and the psychiatric disorder(s)? and (2) which genes are implicated in determining the development of specific symptoms and/or psychiatric diagnosis once a biological alteration (e.g. hypothalamic-pituitary-adrenal (HPA)-axis dysregulation) has been brought about?

How to model the 'G' component and manage the risk for spurious findings

Researchers in genetics have learned that, in order to reduce the risk for spurious findings, hypotheses

should be tested in a systematic manner, with attention to stringent control for multiple testing and a focus on efforts at (exact) replication. There are no reasons to believe that these recommendations do not apply for the G×E field. It is, however, important to realise that the challenges in the field of G×E are greater, given the large number of possible interactions between different environmental factors and the genome, and the fact that available sample sizes are likely to be smaller in G×E studies given the need for extensive phenotyping.

In the conceptual framework of aetiological stratification; however, gene-finding is not a goal in itself. Rather, the goal is to identify subgroups of patients characterised by similar aetiological mechanisms, phenotypic characteristics and functional outcome, who may benefit from treatments informed by the underlying aetiology. In this framework, GWA studies may serve as a useful resource that can be used to derive informed decisions for the appropriate choice of genetic markers. These may include individual markers for which a strong rationale is available (e.g. genome-wide significant association with a disorder or underlying trait), polygenic risk scores, based on large GWA studies of disorders (e.g. schizophrenia) or underlying traits (e.g. neuroticism) but may also involve biological pathways known to be implicated in the biological effects of the environmental factor under examination. That is, while GWA studies may be a useful resource for gene selection, they should not be considered the only source.

In general, it should be considered that GWA studies have shown that many SNPs with individually small effects are likely to be implicated in psychiatric disorders (typically individual SNPs have odds ratios (ORs) under 1.20) (Sullivan *et al.* 2012). Given that most environmental exposures relevant to psychiatric disorders are quite prevalent in the population, there are either large and systematic advantages to carrying 'risk alleles' in the absence of exposure (i.e. they are 'plasticity' genes that promote adaptation to the environment in advantageous conditions but increase risk for maladaptation to non-advantageous conditions), or a recalibration of anticipated effect sizes of individual SNPs in G×E research towards more moderate effects (e.g. ORs of about 2 rather than ORs of 10) may be needed. Given that the prevalence of exposure to different environmental factors varies considerably from sample to sample in GWA studies, but reported ORs do not, the latter possibility seems most likely.

This has important implications for G×E research. As the goal of G×E research is not gene hunting but is situated in the identification of biological mechanisms in subgroups exposed to the same aetiological factor, the examination of genetic measures at the

aggregate level, for example in biological pathways or polygenic risk scores, may be more informative than the examination of individual SNPs. This approach also reduces the problem of multiple testing and the associated risk of spurious findings.

Conclusion

In this editorial, I have tried to outline that G×E research is not an alternative method for GWA studies to find genes for psychiatric disorders, but a way to identify genetic and biological mechanisms within subgroups of patients exposed to a similar aetiological factor. This is important as current diagnostic categories are inherently heterogeneous and lack biological validity. The study of subgroups of patients stratified for exposure to important (environmental) aetiological factors, within and across diagnostic categories, may help to resolve this problem. Careful examination of epidemiological patterns associated with relevant exposures, combined with efforts to identify specific phenotypic or molecular characteristics in exposed subgroups, can be helpful in making the appropriate decisions about the selection of genetic measures and phenotypes to examine. This framework resonates with calls for the application of ‘stratified medicine’ approaches in psychiatry, opening the possibility for the application of personalised medicine approaches in the area of mental health.

Financial Support

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Statement of Interest

None.

References

- Caspi A, Hariri A, Holmes A, Uher R, Moffitt TE (2010). Genetic sensitivity to the environment: the case of the serotonin transporter gene and its implications for studying complex diseases and traits. *American Journal of Psychiatry* 167, 509–527.
- Chun S, Westmoreland JJ, Bayazitov IT, Eddins D, Pani AK, Smeyne RJ, Yu J, Blundon JA, Zakharenko SS (2014). Specific disruption of thalamic inputs to the auditory cortex in schizophrenia models. *Science* 344, 1178–1182.
- Cross-Disorder Group of the Psychiatric Genomics Consortium, Smoller JW, Craddock N, Kendler KS, Lee PH, Neale B, Nurnberger JL, Ripke S, Santangelo S,

Sullivan P (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381, 1371–1379.

Duncan LE, Keller MC (2011). A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry* 168, 1041–1049.

Howes O, Kapur S (2009). The dopamine hypothesis of schizophrenia: version III – the final common pathway. *Schizophrenia Bulletin* 35, 549–562.

Klengel T, Mehta D, Anacker C, Rex-Haffner M, Pruessner J, Pariante CM, Pace T, Mercer KB, Mayberg HS, Bradley B, Nemeroff CB, Holsboer F, Heim CM, Ressler KJ, Rein T, Binder EB (2013). Allele-specific FKBP5 DNA demethylation mediates gene-childhood trauma interactions. *Nature Neuroscience* 16, 33–41.

Mehta D, Klengel T, Conneely KN, Smith AK, Altmann A, Pace T, Rex-Haffner M, Loeschner A, Gonik M, Mercer KB, Bradley B, Muller-Myhsok B, Ressler KJ, Binder EB (2013). Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proceedings of the National Academy of Sciences of the United States of America* 110, 8302–8307.

Modinos G, Iyegbe C, Prata D, Rivera M, Kempton MJ, Valmaggia LR, Sham P, van Os J, McGuire P (2013). Molecular genetic gene-environment studies using candidate genes in schizophrenia: a systematic review. *Schizophrenia Research* 150, 356–365.

Moore TH, Zammit S, Lingford-Hughes A, Barnes TR, Jones PB, Burke M, Lewis G (2007). Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *Lancet* 370, 319–328.

Ogino S, Chan AT, Fuchs CS, Giovannucci E (2011). Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. *Gut* 60, 397–411.

Ogino S, Lochhead P, Chan AT, Nishihara R, Cho E, Wolpin B, Meyerhardt J, Meissner A, Schernhammer E, Fuchs CS, Giovannucci E (2013). Molecular pathological epidemiology of epigenetics: emerging integrative science to analyze environment, host, and disease. *Modern Pathology* 26, 465–484.

Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P, Purcell Leader SM, Stone JL, Sullivan PF, Ruderfer DM, McQuillin A, Morris DW, O'Dushlaine CT, Corvin A, Holmans PA, O'Donovan MC, Sklar P, Wray NR, Macgregor S, Sklar P, Sullivan PF, O'Donovan MC, Visscher PM, Gurling H, Blackwood DH, Corvin A, Craddock NJ, Gill M, Hultman CM, Kirov GK, Lichtenstein P, McQuillin A, Muir WJ, O'Donovan MC, Owen MJ, Pato CN, Purcell SM, Scolnick EM, St Clair D, Stone JL, Sullivan PF, Sklar Leader P, O'Donovan MC, Kirov GK, Craddock NJ, Holmans PA, Williams NM, Georgieva L, Nikolov I, Norton N, Williams H, Toncheva D, Milanova V, Owen MJ, Hultman CM, Lichtenstein P, Thelander EF, Sullivan P, Morris DW, O'Dushlaine CT, Kenny E, Quinn EM, Gill M, Corvin A, McQuillin A, Choudhury K, Datta S, Pimm J, Thirumalai S, Puri V, Krasucki R, Lawrence J, Quesed

- D, Bass N, Gurling H, Crombie C, Fraser G, Leh Kuan S, Walker N, St Clair D, Blackwood DH, Muir WJ, McGhee KA, Pickard B, Malloy P, Maclean AW, Van Beck M, Wray NR, Macgregor S, Visscher PM, Pato MT, Medeiros H, Middleton F, Carvalho C, Morley C, Fanous A, Conti D, Knowles JA, Paz Ferreira C, Macedo A, Helena Azevedo M, Pato CN, Stone JL, Ruderfer DM, Kirby AN, Ferreira MA, Daly MJ, Purcell SM, Sklar P, Purcell SM, Stone JL, Chambert K, Ruderfer DM, Kuruvilla F, Gabriel SB, Ardlie K, Moran JL, Daly MJ, Scolnick EM, Sklar P (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752.
- Sebat J, Levy DL, McCarthy SE (2009). Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends in Genetics* **25**, 528–535.
- Sullivan P, Daly MJ, O'Donovan MC (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nature Review Genetics* **13**, 537–551.
- Sullivan PF (2008). Schizophrenia genetics: the search for a hard lead. *Current Opinion in Psychiatry* **21**, 157–160.
- van Nierop M, Bak M, De Graaf R, Ten Have M, van Dorsselaer S, Genetic Risk and Outcome of Psychosis (GROUP) Investigators, van Winkel R (2014a). Evidence for the social and clinical relevance of childhood trauma-related admixture of affective, anxious and psychosis symptoms, submitted.
- van Nierop M, Viechtbauer W, Gunther N, van Zelst C, De Graaf R, Ten Have M, van Dorsselaer S, Bak M, Genetic Risk and Outcome of Psychosis (GROUP) Investigators, van Winkel R (2014b). Childhood trauma is associated with a specific admixture of affective, anxiety, and psychosis symptoms cutting across traditional diagnostic boundaries. *Psychological Medicine*, revised version currently under review.
- van Winkel R, Kuepper R (2014). Epidemiological, neurobiological, and genetic clues to the mechanisms linking cannabis use to risk for nonaffective psychosis. *Annual Review of Clinical Psychology* **10**, 767–791.
- van Winkel R, van Nierop M, Myin-Germeys I, van Os J (2013). Childhood trauma as a cause of psychosis: linking genes, psychology, and biology. *Canadian Journal of Psychiatry* **58**, 44–51.
- Zammit S, Owen M, Lewis G (2010). Misconceptions about gene-environment interactions in psychiatry. *Evidence Based Mental Health* **13**, 65–68.