Title

Boosting the Power of Schizophrenia Genetics by Leveraging New Statistical Tools

Journal

Schizophrenia Bulletin, 40(1)

ISSN

0586-7614

Authors

Andreassen, Ole A
Thompson, Wesley K
Dale, Anders M

Publication Date

2014

DOI

10.1093/schbul/sbt168

Peer reviewed

Schizophrenia Bulletin vol. 40 no. 1 pp. 13–17, 2014
doi:10.1093/schbul/sbt168
Advance Access publication December 6, 2013

Boosting the Power of Schizophrenia Genetics by Leveraging New Statistical Tools

Ole A. Andreassen*[,1,2], Wesley K. Thompson[3], and Anders M. Dale[3–6]

[1]NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [2]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; [3]Department of Psychiatry, University of California, La Jolla, San Diego, CA; [4]Multimodal Imaging Laboratory, University of California, La Jolla, San Diego, CA; [5]Department of Radiology, University of California, La Jolla, San Diego, CA; [6]Department of Neurosciences, University of California, La Jolla, San Diego, CA

*To whom correspondence should be addressed; NORMENT, KG Jebsen Centre for Psychosis Research, Oslo University Hospital, Building 49, Ullevål, Kirkeveien 166, PO Box 4956, Nydalen, 0424 Oslo, Norway; tel: +47 23 02 73 50 (22 11 78 43 dir), fax: +47 23 02 73 33, e-mail: o.a.andreassen@medisin.uio.no

Genome-wide association studies (GWAS) have identified a large number of gene variants associated with schizophrenia, but these variants explain only a small portion of the heritability. It is becoming increasingly clear that schizophrenia is influenced by many genes, most of which have effects too small to be identified using traditional GWAS statistical methods. By applying recently developed Empirical Bayes statistical approaches, we have demonstrated that functional genic elements show differential contribution to phenotypic variance, with some elements (regulatory regions and exons) showing strong enrichment for association with schizophrenia. Applying related methods, we also showed abundant genetic overlap (pleiotropy) between schizophrenia and other phenotypes, including bipolar disorder, cardiovascular disease risk factors, and multiple sclerosis. We estimated the number of gene variants with effects in schizophrenia and bipolar disorder to be approximately 1.2%. By applying our novel statistical framework, we dramatically improved gene discovery and detected a large number of new gene loci associated with schizophrenia that have not yet been identified with standard GWAS methods. Utilizing independent schizophrenia substudies, we showed that these new loci have high replication rates in de novo samples, indicating that they likely represent true schizophrenia risk genes. The new statistical tools provide a powerful approach for uncovering more of the missing heritability of schizophrenia and other complex disorders. In conclusion, the highly polygenic architecture of schizophrenia strongly suggests the utility of research approaches that recognize schizophrenia neuropathology as a complex dynamic system, with many small gene effects integrated in functional networks.

Key words: GWAS/polygenicity/pleiotropy/empirical Bayes approach/molecular genetics

Introduction

Complex disorders such as schizophrenia (SCZ) are multifactorial and associated with the effects of multiple genes in combination with environmental factors. These disorders often cluster in families, have no clear-cut pattern of inheritance, and have a high fraction of phenotypic variance attributable to genetic variance (high heritability). It is becoming increasingly clear that many genes influence most complex disorders,[1] including SCZ.[2] In such a scenario with a very high number of risk genes ("polygenic"), each gene has a tiny effect.[3] This makes it difficult to determine an individual's risk and to identify disease mechanisms that can be used for development of new effective treatments.

Genome-wide association studies (GWAS) have identified many trait-associated single nucleotide polymorphisms (SNPs),[4] but so far, these explain only small portions of the heritability of complex disorders.[5] SCZ is highly heritable (0.6–0.8), but only a very small fraction of genetic variance has been identified despite recent large, successful GWAS.[2,6–8] This "missing heritability" has been attributed to a number of potential causes, including lack of typing of rare variants.[5] However, it has been shown that a large proportion of the missing heritability is available within GWAS data when associations of SNPs are examined in aggregate.[9] This implies the existence of numerous common variants with small genetic (polygenic) effects. These effects cannot be reliably detected with traditional GWAS statistical methods given current sample sizes. Thus, there is a need for innovative statistical approaches to identify polygenetic effects and reduce the proportion of missing heritability. Utilizing novel statistical approaches can speed discovery more cost efficiently than just collecting larger and larger samples.

In this article, we will describe our recently developed statistical tools that are specifically designed for polygenic disorders, building on an empirical Bayesian framework.[10] These methods both enhance gene discovery and improve replication rates of discovered risk gene variants. This analytical approach is particularly useful for SCZ because of its high heritability, large missing heritability, and evidence that it is highly polygenic[1,2]—a situation that implies that the majority of genetic effects are too small to be identified in traditional GWAS analyses.

Overview of Statistical Methods

Here, we briefly describe our statistical approach for uncovering more of the missing heritability of complex phenotypes using existing GWAS data. These methods have been described in detail in a series of studies investigating psychiatric[11–13] and nonpsychiatric disorders.[13,14]

Q-Q Plots and False Discovery Rates

Q-Q plots are standard tools for assessing similarity or differences between two cumulative distribution functions (CDFs). When the probability distribution of GWAS summary statistic two-tailed P values is of interest, under the global null hypothesis, the theoretical distribution is uniform on the interval [0,1]. If nominal P values are ordered from smallest to largest, so that $P_{(1)} < P_{(2)} < \ldots < P_{(N)}$, the corresponding empirical CDF, denoted by "Q," is simply $Q_{(i)} = i/N$ (in practice, adjusted slightly to account for the discreteness of the empirical CDF), where N is the number of SNPs in the GWAS (or genic category). Thus, for a given index i, the x-coordinate of the Q-Q curve is $Q_{(i)}$ (since the theoretical inverse CDF is the identity function) and the y-coordinate is the nominal P value $P_{(i)}$. It is a common practice in GWAS to instead plot $-\log_{10} P$ against the $-\log_{10} Q$ to emphasize tail probabilities of the theoretical and empirical distributions. For a given threshold of genomic control-corrected P values, "enrichment" is seen as a horizontal deflection of the Q-Q curves from the identity line.

Enrichment seen in the Q-Q plots can be directly interpreted in terms of false discovery rate (FDR). For a given P value cutoff, the Bayes FDR, defined as the posterior probability of a given SNP is null, given its observed P value, is given by:

$$\mathrm{FDR}(P) = \pi_0 F_0(P)/F(P), \tag{1}$$

where π_0 is the proportion of null SNPs, F_0 is the CDF under the null hypothesis, and F is the CDF of all SNPs, both null and non-null. Here, F_0 is the CDF of the uniform distribution on the unit interval [0,1], and $F(P)$ can be estimated with the empirical CDF Q, so that an estimate of equation (1) is given by:

$$\mathrm{FDR}(P) \approx \pi_0 \cdot P/Q, \tag{2}$$

which is biased upwards as an estimate of the FDR. Setting $\pi_0 = 1$ in equation (2), an estimated FDR is further biased upward; if π_0 is close to 1, as is likely true for most GWAS, the increase in bias from equation (2) is minimal. The quantity $1 - P/Q$ is, therefore, biased downward, and hence a conservative estimate of the true discovery rate (equal to 1 FDR). Given the $-\log_{10}$ of the Q-Q plots, we can easily read off:

$$-\log_{10}\big(\mathrm{FDR}(P)\big) \approx \log_{10}(Q) - \log_{10}(P), \tag{3}$$

demonstrating that the (conservatively) estimated FDR is directly related to the horizontal shift of the curves in the Q-Q plots from the expected line $x = y$, with a larger leftward shift corresponding to a smaller FDR.

Conditional Q-Q Plots and FDR. We define the *conditional FDR* as the posterior probability that a SNP belonging to a category c is null for a phenotype, given a P value as small as the observed P value. Formally, this is given by:

$$\mathrm{FDR}(P \mid \mathrm{c}) = \pi_0(\mathrm{c}) \cdot P / F(P \mid \mathrm{c}), \tag{4}$$

where P is the P value for the phenotype, c = 1,…,C is one of C possible categories, $F(P \mid \mathrm{c})$ is the conditional CDF, and $\pi_0(\mathrm{c})$ is the proportion of null SNPs in category c. We produce a conservative estimate of $\mathrm{FDR}(P \mid \mathrm{c})$ by setting $\pi_0(\mathrm{c}) = 1$ and using the empirical conditional CDF in place of $F(P_1 \mid \mathrm{c})$ in equation (4). This is a straightforward generalization of the empirical Bayes approach developed by Efron.[10]

In terms of Q-Q plots, enrichment of category c_2 compared with category c_1 is expressed as a leftward deflection of the Q-Q curve for category c_2 compared with c_1. Given equation (3), this is equivalent to showing that the conditional FDR is smaller for SNPs in category c_2 compared with c_1 for the same P value, ie, $\mathrm{FDR}(P \mid \mathrm{c}_2) < \mathrm{FDR}(P \mid \mathrm{c}_1)$. Thus, by choosing a priori categories that result in differentially enriched samples, a larger proportion of SNPs can be discovered for a given FDR threshold than can be obtained from typical (unconditional) FDR or P value–based analyses.

Covariate-Modulated FDR

We have recently developed a methodology that can capture the combined enrichment signals from several genomic factors, including 5′ untranslated region (UTR), exons, 3′UTR as well as minor allele frequency. Specifically, we are able to incorporate genic annotations and improve gene discovery with a covariate-modulated local FDR (CMFDR).[15] This method includes a data-driven re-ranking of SNPs based on genic annotations, and we have shown that this re-ranking increases yield (number of loci declared non-null) for a given empirical replication rate.

Gene Discovery due to Genomic Enrichment

Using summary statistics derived from SNP associations of huge GWAS, we showed that functional genic elements show differential contribution to phenotypic variance, with some categories (eg, regulatory elements and exons) showing strong *enrichment* (ie, more likely to have an effect) for phenotypic association.[13] The enrichment of SNPs in genic elements of the genome (the 5′UTR and 3′UTR regions) was present across a wide spectrum of complex phenotypes and traits, including SCZ.[13] This shows that SNPs in 5′UTR, in particular, but also in exons and 3′UTR regions are more likely to be involved in susceptibility to SCZ. Although the mechanistic implications of this discovery need to be followed up with experimental studies, this information can be used in Bayesian statistical models to enhance gene discovery by including information on the genic region in which each SNP is located, as this indicates how likely it is for each SNP to have an effect. By applying this approach to data from the Psychiatric Genomics Consortium (PGC) SCZ sample,[16] we substantially increased the power for detecting small genetic effects, leading to discovery of new susceptibility loci that did not reach threshold of significance in traditional GWAS analyses.[13]

Empirical independent replication remains the gold standard for confirming statistical findings. We tested the replication rates, defined as proportion of SNPs declared significant in training samples with *P* values below a given threshold in the replication sample and with *z*-scores with the same sign in both discovery and replication samples, in independent SCZ substudies from the PGC[17] and found that annotation categories with the greatest enrichment (5′UTR, exons, 3′UTR) showed the highest replication rate for a given nominal *P* value, confirming that the observed enrichment is due to true associations and not to inflation due to population stratification or other potential sources of spurious effects (figure 1). These results are all based on summary statistics (*P* values, *z*-scores) for each substudy.

In order to illustrate the increased sensitivity and specificity for gene discovery, we obtained the publically available PGC SCZ sample.[16] Applying the CMFDR method to the PGC SCZ sample, we identified a total of 86 gene loci (CMFDR < 0.05). The gene loci are listed in online supplementary table 1. By computing a posteriori effect sizes from the CMFDR model, we expect that a very large proportion of these loci will replicate in a SCZ GWAS of similar size. With the emerging results from the second phase of the PGC SCZ working group, we will soon know if this prediction is true.

Gene Discovery due to Pleiotropy Enrichment

The small number of genes relative to the vast number of human phenotypes necessitates pleiotropy—the influence of one gene or haplotype on two or more distinct phenotypes. The value of pleiotropy for improved understanding

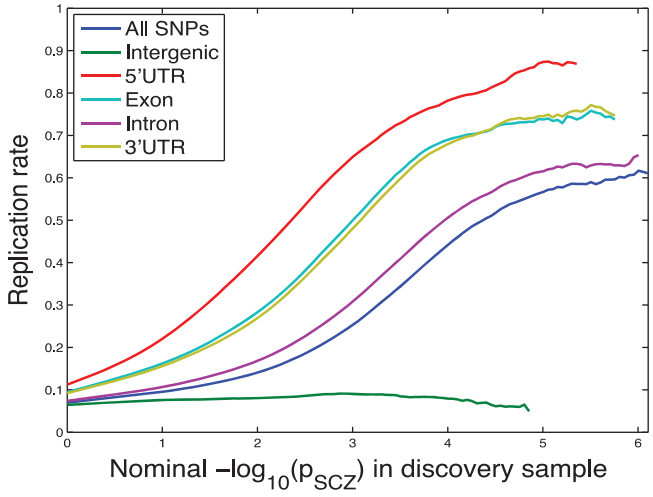

Fig. 1. Cumulative replication plot, showing the average replication rate (y-axis), defined as $P < .05$ in the replication sample and the same sign in both discovery and replication samples, for schizophrenia (SCZ) substudies, for a range of discovery *P* value thresholds (x-axis). SNP, single nucleotide polymorphism; UTR, untranslated region.

of disease pathogenesis and classification, identification of new molecular targets for drug development, and genetic risk profiling have been recognized.[18] But few studies have systematically investigated pleiotropy in human complex traits and disorders, and those that have have looked for pleiotropy only among SNPs that reach a threshold level of significance in one or both phenotypes.[18] This approach fails to capitalize on the power inherent in pleiotropy to robustly detect weak genetic effects.

We applied our novel pleiotropy approach to assess the contribution of all SNPs from two independent GWAS to determine their common association with two distinct phenotypes. SCZ and bipolar disorder share several clinical phenotypes, and there is growing evidence indicating overlapping gene variants.[6,16] We used this approach to increase gene discovery in these disorders, using two large GWAS from the PGC,[6,16] where overlapping controls had been removed with same procedure as in the recent cross-disorder analysis.[19] We discovered a very high degree of polygenic overlap between SCZ and bipolar disorder.[12] We then used this information to increase the power of the GWAS, by including level of pleiotropy as a factor in the statistical models. This resulted in an improved yield (sensitivity) of genes discovered for SCZ and bipolar disorder compared to standard methods at a given significance level (specificity).[12] Thus, by applying the pleiotropy enrichment method and leveraging the bipolar disorder GWAS, we increased gene discovery in the SCZ GWAS. Note, while the power to detect nonpleiotropic loci is not increased using the pleiotropy enrichment method, neither is power lost.

Simulations showed that a larger increase in gene discovery would occur, using standard GWAS approaches, if the SCZ sample was as large as the combined SCZ and

bipolar disorder GWAS.[12] However, it is very expensive to recruit and genotype new samples; applying the new statistical tools to existing samples is a cost-efficient way to improve gene discovery.

Our results also showed that an estimated 1.2% of all SNPs analyzed are pleiotropic for SCZ and bipolar disorder. With approximately 1 million SNPs analyzed, this means that there are approximately 12 000 SNPs involved. This is very similar to the estimate from a recent large SCZ GWAS.[7] This quantification of the polygenicity further emphasizes that most of these variants must have very small effects.

The new statistical tools can also be used to investigate genetic overlap between SCZ and nonpsychiatric diseases and traits to gain more knowledge about shared genetic mechanisms. There is a well-known comorbidity between SCZ and cardiovascular risk factors, including obesity, hypertension, and dyslipidemia.[20] For each of these phenotypes, results are available from large GWAS. We used our pleiotropy methods to investigate polygenic pleiotropy, and found a striking genetic overlap between SCZ and several cardiovascular risk factors, particularly blood lipids (cholesterol, triglycerides). We leveraged this enrichment to boost gene discovery and identified several gene loci associated with SCZ,[11] strongly suggesting that common molecular genetic mechanisms are underlying some of the epidemiological relationships between SCZ and cardiovascular risk factors.

Immune factors have been implicated in SCZ. By investigating pleiotropy with multiple sclerosis, a demyelination disorder with clear evidence for involvement of immune genes, we applied the new statistical tools to determine polygenic overlap. We found a strong genetic overlap between SCZ and multiple sclerosis[21] and identified several independent loci associated with SCZ. In contrast, we found no genetic overlap between bipolar disorder and multiple sclerosis. Imputation of the major histocompatibility complex (MHC) alleles indicated opposite direction of effect in multiple sclerosis and SCZ. As most of the overlap between multiple sclerosis and SCZ was located in the MHC region, and there is previous evidence for large genetic overlap between bipolar disorder and SCZ, our findings suggest that the MHC region could differentiate between bipolar disorder and SCZ.

Polygenic Architecture: Implications for Disease Mechanisms and Clinic

The underlying biology of complex brain disorders such as SCZ remains mostly unknown. Structural magnetic resonance imaging (MRI) brain phenotypes are highly heritable (80%–90%),[22] and a new cluster analytical method has shown how pleiotropic brain phenotypes cluster together.[17] Previous work has shown how a selected number of SNPs can be used to identify genetically determined brain structure variation.[23,24] Recent large meta-analysis showed how brain structure volumes can be successfully used in a GWAS, and SNPs associated with hippocampal volume were identified.[25] By extending a twin study–based approach to a large MRI sample across different behavioral phenotypes, combined with our new statistical framework for analysis of GWAS data to identify polygenic effects, we expect to be able to identify genetically determined brain substrates related to SCZ and core disease phenotypes.

It remains a challenge to translate the new knowledge about the polygenic architecture of SCZ into disease mechanisms. In a scenario with hundreds of genes involved, each with a small effect, it will be an enormous endeavor to extensively characterize the functional consequences of individual gene variants. In addition, it would probably also be important to elucidate the combined effects of the gene variants, which is difficult to test in standard techniques, such as transgenic animal models. Thus, SCZ pathophysiology may be best studied as a complex dynamic system, with a range of interacting small gene effects integrated in functional networks. One obvious aim would be to determine if there are converging functional consequences related to pathophysiological mechanisms already implicated in SCZ, such as dopamine neurotransmission, neurodevelopment, or synaptic function.[26] Such an approach will only be made possible by improved biostatistical tools and computational models. Further, recent evidence also suggests that stem cell technologies is promising for studies of human brain development[27] and may be a new experimental approach to elucidate some of the functional implications of the polygenic architecture of SCZ.

A critical challenge in genetics is to be able to generalize phenotypic predictions and explained variance from existing samples into future datasets. Recently, Purcell et al[2] demonstrated one method for assigning polygenic risk scores by selecting SNPs according to significance thresholds. We propose to leverage our new discoveries of genomic annotations and pleiotropy and their relative enrichments to better select plausible candidate SNPs for calculating and testing generalization performance of polygenic risk scores. Further, we will leverage more powerful statistical techniques by building on empirical Bayesian mixture models, and preliminary results indicate improved prediction based on genomic annotation and polygenic enrichment. The new methods in combination with increasing SCZ GWAS sizes have the potential to produce a number of new insights into the genetic etiology of SCZ, and this could lead to development of personalized medicine approaches and individual prediction of disease risk.

Supplementary Material

Supplementary material is available at http://schizophreniabulletin.oxfordjournals.org.

Funding

O.A.A. was supported by the Research Council of Norway (213837, 217776, 223273), the South East Norway Health Authority (2010-074), and The KG Jebsen Foundation. A.M.D. was supported by a grant from the National Institutes of Health (NIH) (T32 EB005970). W.K.T. was supported by NIH (RC2DA029475 and R01HD061414).

Acknowledgments

The authors would like to thank Linda McEvoy and Yunpeng Wang for helpful comments on the manuscript, our partners in the Psychiatric Genomics Consortium for access to GWAS results from schizophrenia and bipolar disorder, and other GWAS consortium for providing summary statistics data for the individual studies included in this review. The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

1. Wagner GP, Zhang J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat Rev Genet*. 2011;12:204–213.
2. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.
3. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. *Science*. 2002;298:2345–2349.
4. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106:9362–9367.
5. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–753.
6. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium; Ripke S, Sanders AR, Kendler KS, et al. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011;43:969–976.
7. Ripke S, O'Dushlaine C, Chambert K, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet*. 2013;45:1150–1159.
8. Stefansson H, Ophoff RA, Steinberg S, et al. Genetic Risk and Outcome in Psychosis (GROUP). Common variants conferring risk of schizophrenia. *Nature*. 2009;460:744–747.
9. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565–569.
10. Efron B. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge, UK: Cambridge University Press; 2010.
11. Andreassen OA, Djurovic S, Thompson WK, et al. International Consortium for Blood Pressure GWAS; Diabetes Genetics Replication and Meta-analysis Consortium; Psychiatric Genomics Consortium Schizophrenia Working Group. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *Am J Hum Genet*. 2013;92:197–209.
12. Andreassen OA, Thompson WK, Schork AJ, et al. Psychiatric Genomics Consortium (PGC); Bipolar Disorder and Schizophrenia Working Groups. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet*. 2013;9:e1003455.
13. Schork AJ, Thompson WK, Pham P, et al. Tobacco and Genetics Consortium; Bipolar Disorder Psychiatric Genomics Consortium; Schizophrenia Psychiatric Genomics Consortium. All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet*. 2013;9:e1003449.
14. Liu JZ, Hov JR, Folseraas T, et al. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet*. 2013;45:670–675.
15. Zablocki RW, Levine RA, Schork AJ, Andreassen OA, Dale AM, Thompson WK. Covariate-modulated local false discovery rate for genome-wide association studies. *Bioinformatics*. In press.
16. Sklar P, Ripke S, Scott LJ, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet*. 2011;43:977–983.
17. Chen CH, Panizzon MS, Eyler LT, et al. Genetic influences on cortical regionalization in the human brain. *Neuron*. 2011;72:537–544.
18. Sivakumaran S, Agakov F, Theodoratou E, et al. Abundant pleiotropy in human complex diseases and traits. *Am J Hum Genet*. 2011;89:607–618.
19. Cross-Disorder Group of the Psychiatric Genomics Consortium; Genetic Risk Outcome of Psychosis (GROUP) Consortium, Smoller JW, Ripke S, Lee PH, et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. 2013;381:1371–1379.
20. Birkenaes AB, Opjordsmoen S, Brunborg C, et al. The level of cardiovascular risk factors in bipolar disorder equals that of schizophrenia: a comparative study. *J Clin Psychiatry*. 2007;68:917–923.
21. Andreassen OA, Harbo HF, Wang Y, et al. Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune related gene loci. *Mol Psychiatry*. In press.
22. Panizzon MS, Fennema-Notestine C, Eyler LT, et al. Distinct genetic influences on cortical surface area and cortical thickness. *Cereb Cortex*. 2009;19:2728–2735.
23. Joyner AH, J CR, Bloss CS, et al. A common MECP2 haplotype associates with reduced cortical surface area in humans in two independent populations. *Proc Natl Acad Sci U S A*. 2009;106:15483–15488.
24. Rimol LM, Agartz I, Djurovic S, et al.Alzheimer's Disease Neuroimaging Initiative. Sex-dependent association of common variants of microcephaly genes with brain structure. *Proc Natl Acad Sci U S A*. 2010;107:384–388.
25. Stein JL, Medland SE, Vasquez AA, et al. Alzheimer's Disease Neuroimaging Initiative; EPIGEN Consortium; IMAGEN Consortium; Saguenay Youth Study Group; Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium; Enhancing Neuro Imaging Genetics through Meta-Analysis Consortium. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet*. 2012;44:552–561.
26. van Os J, Kapur S. Schizophrenia. *Lancet*. 2009;374:635–645.
27. Lancaster MA, Renner M, Martin CA, et al. Cerebral organoids model human brain development and microcephaly. *Nature*. 2013;501:373–379.