The genetic architecture of autism spectrum disorders (ASDs) and the potential importance of common regulatory genetic variants

David SAFFEN

Institutes of Brain Science, Fudan University, Shanghai 200032, China

Currently, there is great interest in identifying genetic variants that contribute to the risk of developing autism spectrum disorders (ASDs), due in part to recent increases in the frequency of diagnosis of these disorders worldwide. While there is nearly universal agreement that ASDs are complex diseases, with multiple genetic and environmental contributing factors, there is less agreement concerning the relative importance of common *vs* rare genetic variants in ASD liability. Recent observations that rare mutations and copy number variants (CNVs) are frequently associated with ASDs, combined with reduced fecundity of individuals with these disorders, has led to the hypothesis that ASDs are caused primarily by *de novo* or rare genetic mutations. Based on this model, large-scale whole-genome DNA sequencing has been proposed as the most appropriate method for discovering ASD liability genes. While this approach will undoubtedly identify many novel candidate genes and produce important new insights concerning the genetic causes of these disorders, a full accounting of the genetics of ASDs will be incomplete absent an understanding of the contributions of common regulatory variants, which are likely to influence ASD liability by modifying the effects of rare variants or, by assuming unfavorable combinations, directly produce these disorders. Because it is not yet possible to identify regulatory genetic variants by examination of DNA sequences alone, their identification will require experimentation. In this essay, I discuss these issues and describe the advantages of measurements of allelic expression imbalance (AEI) of mRNA expression for identifying *cis*-acting regulatory variants that contribute to ASDs.

common-disease-common-variant model, common-disease-rare-variant model, copy number variant (CNV), allelic expression imbalance (AEI)

ASDs comprise a group of complex disorders characterized by language impairments, social deficits and repetitive and stereotyped patterns of behaviors [1]. The three primary forms of ASD—autistic disorder, pervasive developmental disorder-not otherwise specified and Asperger's syndrome—are usually diagnosed in early childhood and are widely considered to be caused by abnormalities in neurodevelopment [2]. The prevalence of ASD among 8-year olds in the United States is estimated to be about 9 in 1000 (0.9%), with an approximately 4-fold higher occurrence in males compared to females [3]. The prevalence of ASD in Asia and Europe is reported to be 0.6%–1% [3]. Twin and family studies have shown that ASD are strongly influenced by genetics, with an estimated heritability of 0.9 for autistic disorder [4]. An estimated 7%–10% of ASD cases are strongly associated with "syndromic" disorders caused by mutations in single genes (e.g., *FMR1* in fragile X syndrome [5], *MECP2* in Rett syndrome, and *TSC1* and *TSC2* in tuberous sclerosis complex) or by deletions or duplications affecting multiple genes (e.g., deletion of 16p11.2 [6,7] or maternal duplication/triplication of 15q11-q13 [8,9]). The genetic basis of non-syndromic (i.e., "idiopathic") ASD, however, is poorly understood [10].

Two paradigms for the genetic "architecture" of ASD,

email: saffen@fudan.edu.cn

the common-variant and rare-variant models, respectively, have provided the basis for attempts to experimentally identify ASD liability genes [10]. Until recently, the common-variant model was widely assumed to apply to ASD and provided the rationale for many candidate gene and genome-wide association studies. The relatively small yield of novel ASD susceptibility genes identified in these studies, however, has recently stimulated interest in the rare-variant model. The idea that rare, highly penetrant genetic variants, rather than common variants of modest effect, are the primary of cause of many common disorders, including ASD and schizophrenia, has been forcibly argued in recent reviews [11,12].

The goal of this essay is to briefly describe the strengths and weaknesses of these two models as they apply to ASD and to discuss the probable need to include common regulatory variants, i.e., genetic variants that influence mRNA or protein expression, rather than change the amino acid sequence of the encoded protein, in any accounting of the genetic basis of ASD and the methods by which these variants can be identified.

1   The common-variant model for ASD

The common-variant hypothesis [13] proposes that susceptibility to common disorders depends upon combinations of genetic variants that occur frequently (population frequencies ⩾1%), with some combinations increasing susceptibility and others possibly providing protection against disorders. Individually, disease susceptibility alleles are proposed to make only a modest contribution to disease and, therefore, not be subject to severe selection that would rapidly eliminate them from the population. Mathematical modeling shows that in the absence of strong negative selection, many of the genetic variants that contributed to common disorders in our "out-of-Africa" ancestors would continue to persist in modern populations [14]. Thus, ancient genetic variants found in current human populations can be expected to contribute to current common disorders. Examples of common, and possibly ancient, variants that influence the risk of complex diseases include the apolipoprotein E4 (*APOE-4*) allele in Alzheimer's disease [15,16], the transmembrane conductance regulator-F508 deletion (*CFTR*-F508Δ) allele in cystic fibrosis [17], and the complement factor H (*CFH*)-Y402H allele in age-related macular degeneration [18].

Because individuals with ASD have reduced reproductive fitness, it would seem that there should be severe selection against the genetic variants that cause this disorder. Other observations, however, suggest that ASD-related variants may exist in the larger population under less stringent negative selection. For example, sub-clinical ASD-like traits can be detected in the relatives of "multiplex" families who have two-or-more children with ASD [19] and some autism-associated genetic markers have been shown to corre-

late with social communication phenotypes in the general population [20]. From this perspective, clinically diagnosed ASD can be thought to occupy the extreme region of a spectrum of genetically influenced behaviors that extends over the entire population. Such a model argues in favor of contributions to ASD from multiple common genetic variants. Individuals in the clinical range of the spectrum, could simply be those who have inherited unfortunate combinations of variants, which in isolation do not produce debilitating phenotypes.

Common genetic variants linked to ASD susceptibility have, in fact, been identified in genome-wide association studies. Examples include a single nucleotide polymorphism (SNP) located between the cadherin genes *CDH9* and *CDH10* on chromosome 15p14.1 and a SNP near the semaphorin 5A gene *SEMA5A* on chromosome 5p15 [21–23]. The predicted effect sizes associated with these variants, however, are small and functional variants that change the properties of specific genes have not yet been identified. Stronger evidence for association with ASD has been obtained in family-based linkage and association studies for a common variant (rs7794745) in the neurexin super-family gene *CNTNAP2* [24–26]. The majority of ASD candidate gene-based association studies, however, have produced inconsistent evidence for contributions of common variants [27]. In addition, a recent study has raised the possibility that many statistically significant associations between common variants and phenotypes may actually be "synthetic" associations driven by rare causal variants [28]. The failure of association studies to clearly identify multiple genes that contribute to ASD has been interpreted by many investigators as evidence that common variants do not significantly contribute to these disorders.

2   The rare variant model for ASD

The rare-variant model for common disorders proposes that disease susceptibility depends upon many individually rare (population frequencies < 1%), highly penetrant genetic variants [29–31]. Rare variants include point mutations, small and large deletions and duplications and chromosomal rearrangements that disrupt the structure or function of genes. Especially for disorders with reduced fertility, mutations that cause the disorder are proposed to arise *de novo* in the gametes that produced the affected individual or to be propagated for only a few generations within families. Due to strong negative selection, such mutations would never become "common" variants in the population.

One of the most exciting developments in research on the etiology of ASD has been the recent discovery that many cases of ASD are associated with rare deletions or duplications, termed copy number variants (CNVs), that directly or indirectly affect genes related neurodevelopment and/or synaptic functions [32–35]. Genes affected by CNVs linked

to ASD include *NRXN1* [36], *NLGN4X* [37], *APBA2* [38], *SYNGAP1* [35], *DLGAP2* [35], *SHANK2* [35,39], and *SHANK3* [34]. DNA sequencing has also identified rare missense and nonsense mutations in several of these genes in ASD patients [10].

Although, evidence for the involvement of rare CNVs and other genetic variants in ASD is rapidly accumulating, association with ASD is often based simply on the presence of an excess number of these variants among cases compared to controls, with some variants detected in cases only a few times [32–35]. The observation that many rare CNVs and point mutations associated with ASD are also found in individuals with mental retardation or schizophrenia [40] and, at lower frequencies, in control populations [32–35], suggests that individual variants must interact with other factors to produce specific pathological phenotypes. Although combinations of rare variants could in principal cause ASD, it is statistically more likely that that individual rare variant interact with common variants to produce pathological or protective effects. The following sections discuss reasons why common regulatory genetic variants are likely to contribute to these disorders and methods by which these variants can be detected.

3   The potential importance of regulatory genetic variants in ASD

Genetic variation located outside of protein coding regions is common in the human genome and is thought to play a major role in producing phenotypic differences among individuals, including susceptibility to disease [41–46]. The majority of these variants are likely to be evolutionarily "neutral" and lack physiological effects. Many genetic variants in non-coding DNA regions, however, have been shown to regulate gene expression by modulating mRNA transcription, splicing or stability [47]. Although most regulatory variants are located in close proximity to the genes they influence, this is not always the case. "*Cis*-acting" variants that exert strong influences on mRNA expression (via mechanisms not involving diffusible factors), for example, have been detected at great distances (>10–1000 kb) from the start of transcription [48]. A well-documented example is a variant ($-13910 \times T$) common in European populations located 13 kb upstream from the human lactase gene *LCT* that allows expression of lactase to persist in adults, thereby affording protection against adult-type hypolactasia [49].

Compared to mutations that alter protein structure or gene copy number, regulatory variants can be expected to produce "milder" effects by changing levels of gene expression rather than completely knocking out gene function. In addition to directly contributing to disease susceptibility, common regulatory variants have been shown to modulate the effects of rare pathogenic variants [50,51].

There are several reasons why regulatory variants are

difficult to detect in genetic association studies. In candidate gene-based association studies, for example, *cis*-acting regulatory variants located at a great distance from the candidate gene may not be selected for study. Using mRNA expression as a phenotype, it is often possible to identify distant regulatory variants and employ these as markers for linking gene to phenotype [52].

More generally, pathologic (or protective) effects of regulatory variants may be hard to detect because, individually, they make only small contributions to gene expression. In fact, recent studies have shown that mRNA expression for some genes is a "complex trait", influenced by multiple regulatory elements [53–55]. Especially in the case of large genes, individual regulatory variants may reside within independent haplotype blocks. For these genes, studies looking for statistical associations genetic marker-by-genetic marker may fail to detect contributions of individual variants. Using RNA expression as a phenotype allows the contributions of multiple regulatory elements to be directly assessed [53–55], thereby facilitating the selection of optimal sets of markers for genetic association studies.

Similar arguments hold for regulatory variants that influence genes within the same biological pathway or system: the effects of variants within individual genes may be small, but variants in multiple genes may combine to produce large effects [56]. To deal with these complexities, there are clear advantages to independently identifying genetic variants that, singly or in combination, influence mRNA expression of key genes and using these (or surrogate "indicator" SNPs) as markers in genetic association studies. Specifically, identifying high- and low-expression alleles for key genes may help to predict how combinations of these alleles influence disease susceptibility or protection. For example, a low-expression variant for mRNA encoding an enzyme that performs a rate-limiting step in the biosynthesis of a neurotransmitter might produce a significant phenotypic effect only in individuals who also harbor a high-expression variant for mRNA encoding an enzyme that degrades the neurotransmitter. It is possible that interactions among high-and low-expression alleles account for much of the "missing heritability" of complex diseases [57].

4   Detecting regulatory variants

The discovery of rare CNVs and point mutations that associate with ASD has strengthened the argument that large-scale DNA sequencing will be an efficient method for discovering additional genetic variants that contribute to these disorders. Although this strategy is compelling, and large-scale whole-genome or exome-centered sequencing is likely to produce a wealth of novel candidate genetic variants [58,59], it is not the appropriate method for discovering regulatory variants. Bioinformatic tools designed to identify variants that influence RNA transcription, splicing, and sta-

bility are currently far from exact, making it unlikely that most variants that affect RNA expression can be predicted from genomic DNA sequences alone. Instead, experiments will be required to identify regulatory variants, for example, by quantifying mRNA expression in the physiologically relevant tissue and identifying genetic variants that correlate with high- or low-levels of expression.

Methods used to quantify mRNA expression include Northern blotting, polymerase chain reaction (PCR), and microarrays. Using these techniques to analyze regulatory variants that contribute to ASD, however, is complicated by the need to measure mRNA expression in human brain tissue. Brain samples are usually obtained at autopsy from "normal" individuals, and there is little that the investigator can do to control for variables that may influence mRNA expression and quality, such as postmortem interval, tissue pH, cause of death, medical and drug histories [60]. These "external" factors introduce considerable variability into measurements of mRNA expression and make it difficult to make meaningful comparisons among samples. Studying mRNA expression in brain tissue using "high-throughput" methods such as microarrays is also complicated by the heterogeneity of cell types, the 10-to-20-fold excess of glial cells over neurons, and the restriction of expression of many mRNAs to highly localized nuclei, such as serotonin-related mRNAs in neurons in raphe nuclei of the brain stem.

As an alternative to the above methods, we [61,62] and others [63,64], have used PCR-based, allele-specific methods for quantifying mRNA expression in individual samples of human brain. In these methods, a SNP in the mature mRNA serves as a molecular marker to distinguish mRNAs derived from each of the two alleles present in individuals who are heterozygous for the SNP. After isolation of total RNA from frozen, non-fixed sections of human brain, cDNA is prepared by reverse-transcription using oligo-dT or gene-specific primers. Segments of the resulting cDNA containing the marker SNP are then PCR-amplified using synthetic oligonucleotide primers that flank the marker SNP. Finally, the relative amounts of PCR products derived from each allele are quantified using a primer-extension assay [61–64] or by DNA sequencing [65]. "Allelic expression imbalance" (AEI) ratios obtained from these measurements are subsequently adjusted to account for experimental bias using a correction factor determined by carrying out the assays using genomic DNA as the initial template [61].

These assays are highly reproducible and allow allelic expression ratios to be determined with an error of 20%, or less. The accuracy and reproducibility of these measurements come from that fact that AEI ratios are determined for individual brain samples, where each allele has been exposed to the same "external" factors. Any difference in mRNA expression between alleles is presumed to reflect the influence of *cis*-acting genetic variants that differentially regulate each allele.

Once AEI has been detected, the regulatory variants responsible for the differential mRNA expression or alternatively an "indicator" SNPs (iSNPs) that are tightly linked to the regulatory variant can often be identified through an association analysis with genotyped SNPs in the neighborhood of the gene. An efficient method for identifying indicator SNPs for many genes has been evaluate to SNPs in the region of the gene for correlations between SNP heterozygosity and AEI and lack of correlation between SNP homozygosity and AEI [62]. For SNPs that satisfy these criteria, "high"- and "low"-expression alleles are identified by plotting absolute mRNA expression levels (quantified by real-time PCR) versus the SNP genotypes (AA, Aa, aa) [62]. Indicator SNPs identified by these assays can be used as genetic markers in case-control association studies. Identification of the actual regulatory variant (presumably tightly linked to the iSNP) often requires additional experiments, including resequencing of chromosomal segments in linkage disequilibrium with the iSNP and/or testing the effects of putative regulatory variants on gene expression, RNA transcript splicing or mRNA stability *in vitro* [64,66,67].

5   How to proceed

Genetic markers obtained by the above procedures are ideal for testing the hypothesis that common regulatory variants contribute to ASD. Because identifying regulatory variants is labor-intensive, it is most suitable for use in candidate gene-based studies. Candidate genes may be selected based on biological theories concerning the etiology of the disorder (in the case of ASD, for example, selecting genes related to neurodevelopment or synaptic function) or based on information about specific genes or biological pathways previously identified in genetic association or linkage studies. In addition, genes that are influenced by recurrent or rare CNVs linked to ASD are an obvious choice.

Because deletion or duplication of a gene in a CNV only affects one of a pair of autosomal genes, a "normal" copy of the gene is usually present on the other chromosome. For this reason, CNVs are most likely to produce pathological effects when they affect "dose-sensitive" genes, for which deletions result in "haploinsufficiency" and duplications result in inappropriate over-expression. In either case, the expected change in gene expression is only 50%, down or up. Surveys of mRNA expression human tissues, including brain, show that allele-specific differences in mRNA expression of 1.5–2-fold are common [68,69]. Regulatory variants can therefore be expected to mimic the effects of single-gene deletions and duplications. For example, individuals harboring two "low-expression" alleles of a dose-sensitive autosomal gene, each expressing 50% less than average, might experience the same risk for disease as individuals who harbor a rare CNV that of deletes one copy for the gene.

Relatively modest changes in gene expression such as those described above have, in fact, been shown to have dramatic effects when they occur in dose-sensitive genes. Good examples are rare CNVs that result in duplication of the amyloid precursor protein gene *APP*. Individuals who harbor duplications, and therefore carry three copies of the *APP* gene, develop Alzheimer's disease (AD) in their early fifties, about 30 years earlier than the typical age-of-onset of sporadic AD [70]. Promoter mutations that cause modest increases in *APP* expression have also been shown to result in early onset AD and to occur at a higher frequency than *APP* coding region variants in at least one AD population [71]. Similarly, individuals with triplication of the alpha-synuclein gene *SNCA* develop early-onset Parkinson's disease [72,73]. Together, these examples show that even modest changes in expression of dose-sensitive genes can have a major influence on disease susceptibility and onset.

As mentioned above, in addition to producing pathological effects, regulatory variants are also likely to play important roles in modulating the effects of gene duplication and deletions produced by CNVs. For example, the presence of a "low-expression" allele of an autosomal gene might ameliorate the effects of duplication of the gene on its paired chromosome, while a high-expression allele might have the opposite effect. The same possibility holds for "high-expression" and "low-expression" alleles paired with pathogenic deletion CNVs. The identification of genetic markers that predict high- and low-mRNA expression for genes located within ASD-associated CNVs will be useful for exploring these possibilities.

6   Summary

The momentum is currently swinging towards the rare-variant model for ASD, a movement driven both by the failure of genome-wide association studies to identify common causal genetic variants and the discovery of rare pathogenic CNVs. Although intensive DNA sequencing is likely to identify many novel mutations and CNVs that contribute to ASD, the inability to identify regulatory variants is likely to leave much of the genetic etiology of these disorders unexplained.

We hypothesize that a complete description of the genetics of ASD is likely to require an accounting of the influences of common regulatory variants. We propose that examining allele-specific differences in mRNA expression for genes located within ASD-associated CNVs will prove to be an effective method for discovering these variants.

1   Freitag CM. The genetics of autistic disorders and its clinical relevance: a review of the literature. Mol Psychiatry, 2007, 12: 2–22

2   Walsh CA, Morrow EM, Rubenstein JL. Autism and brain development. Cell, 2008, 135: 396–400

3   Patel DR, Greydanus DE, Calles JL Jr., Pratt HD. Developmental disabilities across the lifespan. Dis Mon, 2010, 56: 304–397

4   Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, Rutter M. Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med, 1995, 25: 63–77

5   Wang LW, Berry-Kravis E, Hagerman RJ. Fragile X: leading the way for targeted treatments in autism. Neurotherapeutics, 2010, 7: 264–274

6   Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, Samundsen E, Stefansson H, Ferreira MAR, Green T , Platt OS, Rudefer DM, Walsh CA, Altshuler D, Chakravarti A, Tanzi RE, Stefansson K, Santangelo SL, Gusella JF, Sklar P, Wu B , Daly MJ. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med, 2008, 358: 667–675

7   Fernandez BA, Roberts W, Chung B, Weksberg R, Meyn S, Szatmari P, Joseph-George AM, MacKay S, Whitten K, Noble, B, Vardy C, Crosbie V, Luscombe S, Tucker E, Turner L, Marshall CR, Scherer SW. Phenotypic spectrum associated with *de novo* and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. J Med Genet, 47: 195–203

8   Veltman MW, Craig EE, Bolton PF. Autism spectrum disorders in Prader-Willi and Angelman syndromes: a systematic review. Psychiatr Genet, 2005, 15: 243–254

9   Hogart A, Wu D, LaSalle JM, Schanen NC. The comorbidity of autism with the genomic disorders of chromosome 15q11.2-q13. Neurobiol Dis, 38: 181–191

10   Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. Nat Rev Genet, 2008, 9: 341–355

11   McClellan J, King MC. Genetic heterogeneity in human disease. Cell, 2010, 141: 210–217

12   McClellan J, King MC. Genomic analysis of mental illness: a changing landscape. Jama, 2010, 303: 2523–2524

13   Lander ES. The new genomics: global views of biology. Science, 1996, 274: 536–539

14   Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet, 2001, 17: 502–510

15   Strittmatter WJ, Saunders AM, Schmechel D, Pericakvance M, Enghild J, Salvesen GS, ROSES AD. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc Natl Acad Sci USA, 1993, 90: 1977–1981

16   Poirier J, Davignon J, Bouthillier D, Kogan S, Bertrand P, Gauthier S. Apolipoprotein E polymorphism and Alzheimer's disease. Lancet, 1993, 342: 697–699

17   Morral N, Bertranpetit J, Estivill X, Nunes V, Casals T, Gimenez J, Reis A, Varonmateeva R, MACEK M, Kalaydjieva L, Angelicheva D, Dancheva R, Romeo G, Russo MP, Garnerone S, Restagno G, Ferrari M, Magnani C, Claustres M, Desgeorges M, Schwart M, Dallapiccola B, Novelli G, Ferec C, Dearce M, Nemeti M, Kere T, Anvre M, Dahl N, Kadasi L. The origin of the major cystic fibrosis mutation (delta F508) in European populations. Nat Genet, 1994, 7: 169–175

18   Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. Science, 2005, 308: 385–389

19   Virkud YV, Todd RD, Abbacchi AM, Zhang Y, Constantino JN. Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. Am J Med Genet B Neuropsychiatr Genet, 2009, 150B: 328–334

20   St Pourcain B, Wang K, Glessner JT, Golding J, Steer C, Ring SM, Skuse DH, Grant SF, Hakonarson H, Davey SG. Association Between a High-Risk Autism Locus on 5p14 and Social Communication Spectrum Phenotypes in the General Population. Am J Psychiatry, 2010, 167: 1364–1372

21   Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS,

Salyakina D, Imielinski M, Bradfield JP, Sleiman PM, Kim CE, Hou C, Frackelton E, Chiavacci R, Takahashi N, Sakurai T, Rappaport E, Lajonchere CM, Munson J, Estes A, Korvatska O, Piven J, Sonnenblick LI, Alvarez Retuerto AI, Herman EI, Dong H, Hutman T, Sigman M, Ozonoff S, Klin A, Owley T, Sweeney JA, Brune CW, Cantor RM, Bernier R, Gilbert JR, Cuccaro ML, McMahon WM, Miller J, State MW, Wassink TH, Coon H, Levy SE, Schultz RT, Nurnberger JI, Haines JL, Sutcliffe JS, Cook EH, Minshew NJ, Buxbaum JD, Dawson G, Grant SF, Geschwind DH, Pericak-Vance MA, Schellenberg GD, Hakonarson H. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature, 2009, 459: 528–533

22 Weiss LA, Arking DE, Gene Discovery Project of Johns Hopkins & the Autism Consortium, Daly MJ, Chakravarti A. A genome-wide linkage and association scan reveals novel loci for autism. Nature, 2009, 461: 802–808

23 Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Sykes N, Pagnamenta AT, Almeida J, Bacchelli E, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Carson AR, Casallo G, Casey J, Chu SH, Cochrane L, Corsello C, Crawford EL, Crossett A, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Melhem NM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Piven J, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Wing K, Wittemeyer K, Wood S, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Betancur C, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Gallagher L, Geschwind DH, Gill M, Haines JL, Miller J, Monaco AP, Nurnberger JI Jr, Paterson AD, Pericak-Vance MA, Schellenberg GD, Scherer SW, Sutcliffe JS, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM, Devlin B, Ennis S, Hallmayer J. A genomewide scan for common alleles affecting risk for autism. Hum Mol Genet, 2010, 19: 4072–4082

24 Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM, Sebat J, Wigler M, Martin CL, Ledbetter DH, Nelson SF, Cantor RM, Geschwind DH. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-suscepti-bility gene. Am J Hum Genet, 2008, 82: 150–159

25 Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M, Rea A, Guy M, Lin S, Cook EH, Chakravarti A. A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. Am J Hum Genet, 2008, 82: 160–164

26 Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, Morgan TM, Chawarska K, Klin A, Ercan-Sencicek AG, Stillman AA, Tanriover G, Abrahams BS, Duvall JA, Robbins EM, Geschwind DH, Biederer T, Gunel M, Lifton RP, State MW. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet, 2008, 82: 165–173

27 Freitag CM, Staal W, Klauck SM, Duketis E, Waltes R. Genetics of autistic disorders: review and clinical implications. Eur Child Adolesc Psychiatry, 2010, 19: 169–178

28 Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. PLoS Biol, 2010, 8: e1000294

29 Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet, 2001, 69: 124–137

30 Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet, 2008, 40: 695–701

31 Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev, 2009, 19: 212–219

32 Autism Genome Project Consortium, Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L, Feuk L, Qian C, Bryson SE, Jones MB, Marshall CR, Scherer SW, Vieland VJ, Bartlett C, Mangin LV, Goedken R, Segre A, Pericak-Vance MA, Cuccaro ML, Gilbert JR, Wright HH, Abramson RK, Betancur C, Bourgeron T, Gillberg C, Leboyer M, Buxbaum JD, Davis KL, Hollander E, Silverman JM, Hallmayer J, Lotspeich L, Sutcliffe JS, Haines JL, Folstein SE, Piven J, Wassink TH, Sheffield V, Geschwind DH, Bucan M, Brown WT, Cantor RM, Constantino JN, Gilliam TC, Herbert M, Lajonchere C, Ledbetter DH, Lese-Martin C, Miller J, Nelson S, Samango-Sprouse CA, Spence S, State M, Tanzi RE, Coon H, Dawson G, Devlin B, Estes A, Flodman P, Klei L, McMahon WM, Minshew N, Munson J, Korvatska E, Rodier PM, Schellenberg GD, Smith M, Spence MA, Stodgell C, Tepper PG, Wijsman EM, Yu CE, Rogé B, Mantoulan C, Wittemeyer K, Poustka A, Felder B, Klauck SM, Schuster C, Poustka F, Bölte S, Feineis-Matthews S, Herbrecht E, Schmötzer G, Tsiantis J, Papanikolaou K, Maestrini E, Bacchelli E, Blasi F, Carone S, Toma C, Van Engeland H, de Jonge M, Kemner C, Koop F, Langemeijer M, Hijmans C, Staal WG, Baird G, Bolton PF, Rutter ML, Weisblatt E, Green J, Aldred C, Wilkinson JA, Pickles A, Le Couteur A, Berney T, McConachie H, Bailey AJ, Francis K, Honeyman G, Hutchinson A, Parr JR, Wallace S, Monaco AP, Barnby G, Kobayashi K, Lamb JA, Sousa I, Sykes N, Cook EH, Guter SJ, Leventhal BL, Salt J, Lord C, Corsello C, Hus V, Weeks DE, Volkmar F, Tauber M, Fombonne E, Shih A, Meyer KJ. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet, 2007, 39: 319–328

33 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimäki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam TC, Ye K, Wigler M. Strong association of *de novo* copy number mutations with autism. Science, 2007, 316: 445–449

34 Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, Thiruvahindrapduram B, Fiebig A, Schreiber S, Friedman J, Ketelaars CE, Vos YJ, Ficicioglu C, Kirkpatrick S, Nicolson R, Sloman L, Summers A, Gibbons CA, Teebi A, Chitayat D, Weksberg R, Thompson A, Vardy C, Crosbie V, Luscombe N, Baatjes R, Zwaigenbaum L, Roberts W, Fernandez B, Szatmari P, Scherer SW. Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet, 2008, 82: 477–488

35 Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BH, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice

JP, Salt J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H,Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Devlin B, Ennis S, Gallagher L, Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI Jr, Paterson AD, Pericak-Vance MA, Schellenberg GD, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM, Scherer SW, Sutcliffe JS, Betancur C. Functional impact of global rare copy number variation in autism spectrum disorders. Nature, 2010, 466: 368–372

36 Ching MS, Shen Y, Tan WH, Jeste SS, Morrow EM, Chen X, Mukaddes NM, Yoo SY, Hanson E, Hundley R, Austin C, Becker RE, Berry GT, Driscoll K, Engle EC, Friedman S, Gusella JF, Hisama FM, Irons MB, Lafiosca T, LeClair E, Miller DT, Neessen M, Picker JD, Rappaport L, Rooney CM, Sarco DP, Stoler JM, Walsh CA, Wolff RR, Zhang T, Nasir RH, Wu BL, Children's Hospital Boston Genotype Phenotype Study Group. Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. Am J Med Genet B Neuropsychiatr Genet, 2010, 153B: 937–947

37 Lawson-Yuen A, Saldivar JS, Sommer S, Picker J. Familial deletion within NLGN4 associated with autism and Tourette syndrome. Eur J Hum Genet, 2008, 16: 614–618

38 Babatz TD, Kumar RA, Sudi J, Dobyns WB, Christian SL. Copy number and sequence variants implicate APBA2 as an autism candidate gene. Autism Res, 2009, 2: 359–364

39 Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, Endris V, Roberts W, Szatmari P, Pinto D, Bonin M, Riess A, Engels H, Sprengel R, Scherer SW, Rappold GA. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat Genet, 2010, 42: 489–491

40 Guilmatre A, Dubourg C, Mosca AL, Legallic S, Goldenberg A, Drouin-Garraud V, Layet V, Rosier A, Briault S, Bonnet-Brilhault F, Laumonnier F, Odent S, Le Vacon G, Joly-Helas G, David V, Bendavid C, Pinoit JM, Henry C, Impallomeni C, Germano E, Tortorella G, Di Rosa G, Barthelemy C, Andres C, Faivre L, Frébourg T, Saugier Veber P, Campion D. Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. Arch Gen Psychiatry, 2009, 66: 947–956

41 Knight JC. Regulatory polymorphisms underlying complex disease traits. J Mol Med, 2005, 83: 97–109

42 Rockman MV, Wray GA. Abundant raw material for cis-regulatory evolution in humans. Mol Biol Evol, 2002, 19: 1991–2004

43 Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. Mapping complex disease traits with global gene expression. Nat Rev Genet, 2009, 10: 184–194

44 Epstein DJ. Cis-regulatory mutations in human disease. Brief Funct Genomic Proteomic, 2009, 8: 310–316

45 Di Rienzo A, Hudson RR. An evolutionary framework for common diseases: the ancestral-susceptibility model. Trends Genet, 2005, 21: 596–601

46 Sadee W. Regulatory polymorphisms in key candidate genes for disease susceptibility and drug response: a mandate for valid genetic biomarkers. Expert Rev Mol Diagn, 2010, 10: 9–11

47 Johnson AD, Zhang Y, Papp AC, Pinsonneault JK, Lim JE, Saffen D, Dai Z, Wang D, Sadée W. Polymorphisms affecting gene transcription and mRNA processing in pharmacogenetic candidate genes: detection through allelic expression imbalance in human target tissues. Pharmacogenet Genomics, 2008, 18: 781–791

48 Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet, 2005, 76: 8–32

49 Lewinsky RH, Jensen TG, Moller J, Stensballe A, Olsen J, Troelsen JT. T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity *in vitro*. Hum

Mol Genet, 2005, 14: 3945–3953

50 Gouya L, Puy H, Robreau AM, Bourgeois M, Lamoril J, Da Silva V, Grandchamp B, Deybach JC. The penetrance of dominant erythropoietic protoporphyria is modulated by expression of wildtype FECH. Nat Genet, 2002, 30: 27–28

51 Gouya L, Martin-Schmitt C, Robreau AM, Austerlitz F, Da Silva V, Brun P, Simonin S, Lyoumi S, Grandchamp B, Beaumont C, Puy H, Deybach JC. Contribution of a common single-nucleotide polymorphism to the genetic predisposition for erythropoietic protoporphyria. Am J Hum Genet, 2006, 78: 2–14

52 Smith RM, Alachkar H, Papp AC, Wang D, Mash DC, Wang JC, Bierut LJ, Sadee W. Nicotinic alpha5 receptor subunit mRNA expression is associated with distant 5' upstream polymorphisms. Eur J Hum Genet, 2011, 19: 76–83

53 Horan M, Millar DS, Hedderich J, Lewis G, Newsway V, Mo N, Fryklund L, Procter AM, Krawczak M, Cooper DN. Human growth hormone 1 (GH1) gene expression: complex haplotype-dependent influence of polymorphic variation in the proximal promoter and locus control region. Hum Mutat, 2003, 21: 408–423

54 Tao H, Cox DR, Frazer KA. Allele-specific KRT1 expression is a complex trait. PLoS Genet, 2006, 2: e93

55 Babbitt CC, Silverman JS, Haygood R, Reininga JM, Rockman MV, Wray GA. Multiple functional variants in cis modulate PDYN expression. Mol Biol Evol, 2010, 27: 465–479

56 Carlborg O, Haley CS. Epistasis: too often neglected in complex trait studies? Nat Rev Genet, 2004, 5: 618–625

57 Maher B. Personal genomes: The case of the missing heritability. Nature, 2008, 456: 18–21

58 Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet, 2010, 11: 415–425

59 Cirulli ET, Singh A, Shianna KV, Ge D, Smith JP, Maia JM, Heinzen EL, Goedert JJ, Goldstein DB, Center for HIV/AIDS Vaccine Immunology (CHAVI). Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. Genome Biol, 2010, 11: R57

60 Chevyreva I, Faull RL, Green CR, Nicholson LF. Assessing RNA quality in postmortem human brain tissue. Exp Mol Pathol, 2008, 84: 71–77

61 Lim JE, Papp A, Pinsonneault J, Sadée W, Saffen D. Allelic expression of serotonin transporter (SERT) mRNA in human pons: lack of correlation with the polymorphism SERTLPR. Mol Psychiatry, 2006, 11: 649–662

62 Lim JE, Pinsonneault J, Sadee W, Saffen D. Tryptophan hydroxylase 2 (TPH2) haplotypes predict levels of TPH2 mRNA expression in human pons. Mol Psychiatry, 2007, 12: 491–501

63 Zhang Y, Wang D, Johnson AD, Papp AC, Sadée W. Allelic expression imbalance of human mu opioid receptor (OPRM1) caused by variant A118G. J Biol Chem, 2005, 280: 32618–32624

64 Zhang Y, Bertolino A, Fazio L, Blasi G, Rampino A, Romano R, Lee ML, Xiao T, Papp A, Wang D, Sadée W. Polymorphisms in human dopamine D2 receptor gene affect gene expression, splicing, and neuronal activity during working memory. Proc Natl Acad Sci USA, 2007, 104: 20552–20557

65 Xu X, Wang H, Zhu M, Sun Y, Tao Y, He Q, Wang J, Chen L, Saffen D. Next-generation DNA sequencing-based assay for measuring allelic expression imbalance (AEI) of candidate neuropsychiatric disorder genes in human brain. BMC Genomics, 2011, 12: 518

66 Johnson AD, Gong Y, Wang D, Langaee TY, Shin J, Cooper-Dehoff RM, Schork NJ, Binkley P, Pepine CJ, Johnson JA, Sadee W. Promoter polymorphisms in ACE (angiotensin I-converting enzyme) associated with clinical outcomes in hypertension. Clin Pharmacol Ther, 2009, 85: 36–44

67 Zhang X, Nicholls PJ, Laje G, Sotnikova TD, Gainetdinov RR, Albert PR, Rajkowska G, Stockmeier CA, Speer MC, Steffens DC, Austin MC, McMahon FJ, Krishnan KR, Garcia-Blanco MA, Caron MG. A functional alternative splicing mutation in human tryptophan hydroxylase-2. Mol Psychiatry, 2010, 16: 1169–1176

68 Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L,

Kaleem M, Leung D, Bryden L, Nath P, Zismann VL, Joshipura K, Huentelman MJ, Hu-Lince D, Coon KD, Craig DW, Pearson JV, Holmans P, Heward CB, Reiman EM, Stephan D, Hardy J. A survey of genetic human cortical gene expression. Nat Genet, 2007, 39: 1494–1499

69   Webster JA, Gibbs JR, Clarke J, Ray M, Zhang W, Holmans P, Rohrer K, Zhao A, Marlowe L, Kaleem M, McCorquodale DS 3rd, Cuello C, Leung D, Bryden L, Nath P, Zismann VL, Joshipura K, Huentelman MJ, Hu-Lince D, Coon KD, Craig DW, Pearson JV, NACC-Neuropathology Group, Heward CB, Reiman EM, Stephan D, Hardy J, Myers AJ. Genetic control of human brain transcript expression in Alzheimer disease. Am J Hum Genet, 2009, 84: 445–458

70   Singleton A, Myers A, Hardy J. The law of mass action applied to neurodegenerative disease: a hypothesis concerning the etiology and path-ogenesis of complex diseases. Hum Mol Genet, 2004, 13: R123–126

71   Brouwers N, Sleegers K, Engelborghs S, Bogaerts V, Serneels S, Kamali K, Corsmit E, De Leenheir E, Martin JJ, De Deyn PP, Van Broeckhoven C, Theuns J. Genetic risk and transcriptional variability of amyloid precursor protein in Alzheimer's disease. Brain, 2006, 129: 2984–2991

72   Bradbury J. Alpha-synuclein gene triplication discovered in Parkinson's disease. Lancet Neurol, 2003, 2: 715

73   Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, Lincoln S, Crawley A, Hanson M, Maraganore D, Adler C, Cookson MR, Muenter M, Baptista M, Miller D, Blancato J, Hardy J, Gwinn-Hardy K. alpha-Synuclein locus triplication causes Parkinson's disease. Science, 2003, 302: 841