Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia
Tom Walsh, *et al.*
Science **320**, 539 (2008);
DOI: 10.1126/science.1155174

Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia

Tom Walsh,[1]* Jon M. McClellan,[2]*† Shane E. McCarthy,[3]* Anjené M. Addington,[4]* Sarah B. Pierce,[1] Greg M. Cooper,[5] Alex S. Nord,[5] Mary Kusenda,[3,6] Dheeraj Malhotra,[3] Abhishek Bhandari,[3] Sunday M. Stray,[1] Caitlin F. Rippey,[5] Patricia Roccanova,[3] Vlad Makarov,[3] B. Lakshmi,[3] Robert L. Findling,[7] Linmarie Sikich,[8] Thomas Stromberg,[4] Barry Merriman,[9] Nitin Gogtay,[4] Philip Butler,[4] Kristen Eckstrand,[4] Laila Noory,[4] Peter Gochman,[4] Robert Long,[4] Zugen Chen,[9] Sean Davis,[10] Carl Baker,[5] Evan E. Eichler,[5] Paul S. Meltzer,[10] Stanley F. Nelson,[9] Andrew B. Singleton,[11] Ming K. Lee,[1] Judith L. Rapoport,[4] Mary-Claire King,[1,5] Jonathan Sebat[3]

Schizophrenia is a devastating neurodevelopmental disorder whose genetic influences remain elusive. We hypothesize that individually rare structural variants contribute to the illness. Microdeletions and microduplications >100 kilobases were identified by microarray comparative genomic hybridization of genomic DNA from 150 individuals with schizophrenia and 268 ancestry-matched controls. All variants were validated by high-resolution platforms. Novel deletions and duplications of genes were present in 5% of controls versus 15% of cases and 20% of young-onset cases, both highly significant differences. The association was independently replicated in patients with childhood-onset schizophrenia as compared with their parents. Mutations in cases disrupted genes disproportionately from signaling networks controlling neurodevelopment, including neuregulin and glutamate pathways. These results suggest that multiple, individually rare mutations altering genes in neurodevelopmental pathways contribute to schizophrenia.

Schizophrenia is a debilitating neuropsychiatric illness with severe individual, family, and societal burdens. The illness typically arises in late adolescence or early adulthood, with prevalence ~1% worldwide. The phenotype is heterogeneous and complex, with multiple genes and environmental exposures likely involved. Family, twin, and adoption studies all support a strong genetic influence, although patterns of inheritance are variable and not consistent with a single monogenic Mendelian trait (*1*).

The present working hypothesis for genetic influences on schizophrenia is the "common disease–common allele" model (*2*), in which the illness is caused by combinations of common alleles, each contributing a modest effect. We propose an alternative model: that some mutations predisposing to schizophrenia are highly penetrant, individually rare, and of recent origin, even specific to single cases or families (*3*).

Genomic microduplications and microdeletions, also known as structural variants or copy-number variants (CNVs), underlie many serious illnesses, including neurological and neurodevelopmental syndromes (*4*). Previously, we found a significantly increased frequency of de novo deletions and duplications in children with autism spectrum disorders (*5*), suggesting that rare structural variants collectively account for a larger proportion of these disorders than previously recognized. Direct methods for detection of microdeletions and microduplications can be used for gene discovery in other psychiatric disorders as well.

The hypothesis that chromosomal mutations may cause schizophrenia is not new. Karyotypic abnormalities have been detected in affected individuals or families (*6*), including events that led to the discovery of potential disease-causing mutations in *DISC1* (*7*), *PDE4B* (*7*), and *NPAS3* (*8*). Deletions at 22q11.2 are also associated with schizophrenia (*9*). However, the ability to detect structural mutations associated with schizophrenia was historically limited to large events identifiable by karyotype. The development of microarray-based methods now enables the detection of much smaller events. These methods have revealed structural mutations involving genes in neurodevelopmental pathways in individuals with schizophrenia, including a deletion disrupting *NRXN1* in affected siblings (*10*), a de novo duplication involving *APBA2* in one patient (*10*), and deletions containing *CNTNAP2* in two unrelated patients (*11*).

If severe mutations leading to schizophrenia are individually rare, then each individual mutation will be responsible for the disorder's occurrence in only one or a few patients. The overall involvement of rare mutations in the illness can be tested statistically by comparing the collective frequency of rare variants in cases with their collective frequency among controls. We examine whether rare structural mutations throughout the genome are more frequent among persons with schizophrenia than among unaffected individuals. We then examine whether genes disrupted by structural mutations in cases are more likely than genes disrupted in controls to encode proteins critical to brain development.

Genome-wide scans for structural variants greater than 100 kb were carried out on genomic DNA of 418 individuals, including 150 persons with schizophrenia or schizoaffective disorder meeting DSM-IV criteria and 268 healthy individuals (controls) aged 35 years or older who were free of signs of neurological or psychiatric illness. Detailed clinical information regarding the cases is provided in the supporting online material [(*12*) and table S1]. Distributions of racial ancestries were the same in cases and controls. Representational oligonucleotide microarray analysis 85K probe microarrays were used for event discovery. Illumina 550K microarrays and NimbleGen 2.1M feature HD2 microarrays were used to validate events and refine genomic breakpoints (*12*). Stringent quality-control criteria were applied to microarray data to obtain reliable measures of intensity, to exclude false positives, and to ensure that ascertainment of structural variants was consistent between cases and controls (*12*). First, frequencies in cases versus controls were compared for the 115 common variants (i.e., with frequency >1%) that passed our quality-control criteria and were verified by an independent method. Frequencies of these common variants did not differ between cases and controls (fig. S1). Second, detection sensitivity of rare variants was measured by simulating structural variants of various sizes at many different genomic locales and screening for them in blind tests. Detection sensitivity varied by event size and locale but was the same in cases and controls (fig. S2). Variants that were ultimately included in this study passed all these quality-control tests [see also (*12*)] and met our criteria for rare variants, defined as those not previously described on the Database of Genomic Variants, version 3, 29 November 2007 update (*13*). In the 418 cases and controls, we identified and validated 53 previously unreported microdeletions and microduplications, of sizes ranging from 100 kb to 15 MB.

Individuals with schizophrenia were more than three times as likely as controls to harbor rare structural variants that deleted or duplicated one or more genes (Table 1, $P = 0.0008$). Cases with onset of psychotic symptoms at age 18 years or younger were more than four times as likely as controls to harbor such variants (Table 1, $P = 0.0001$). In contrast, there was no significant difference in the proportions

[1]Department of Medicine, University of Washington, Seattle, WA 98195, USA. [2]Department of Psychiatry, University of Washington, Seattle, WA 98195, USA. [3]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. [4]Child Psychiatry Branch, National Institute of Mental Health, National Institutes of Health (NIH), Bethesda, MD 20892, USA. [5]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. [6]Graduate Program in Genetics, State University of New York, Stony Brook, NY 11794, USA. [7]Department of Psychiatry, Case Medical Center, Cleveland, OH 44106, USA. [8]Department of Psychiatry, University of North Carolina, Chapel Hill, NC 27599, USA. [9]Department of Human Genetics, University of California, Los Angeles, CA 90095, USA. [10]Cancer Genetics Branch, National Cancer Institute, NIH, Bethesda, MD 20892, USA. [11]Neurogenetics Laboratory, National Institute on Aging, NIH, Bethesda, MD 20892, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: drjack@u.washington.edu

of cases versus controls carrying rare mutations that did not alter any genes (table S2).

We tested the same hypothesis in an independent series of cases and controls, using a rigorously characterized cohort of youth with childhood-onset schizophrenia (COS). COS is a rare and typically more severe form of the illness (*14*). Of the 92 patients in this cohort, 9 patients known to harbor major chromosomal abnormalities (*15*–*17*) were excluded from this analysis, leaving 83 individuals. Given the heterogeneous ancestries of the COS patients, we evaluated the nontransmitted chromosomes of available parents ($n = 154$) as controls. Parental relationships were validated by genotyping multiple polymorphic markers. Because each parent contributed one haploid nontransmitted genome to analysis, the effective diploid sample size of our control group was 154/2 or 77. Genome-wide scans for structural variants >100 kb and ≥20 single-nucleotide polymorphisms (SNPs) were carried out on genomic DNA from COS cases and their parents by Affymetrix 500K SNP arrays (Nsp I and Sty I chips). All events were validated by Agilent 185K or 244K array comparative genomic hybridization and by custom bacterial artificial chromosome arrays targeting genomic regions of segmental duplications (*18*). Rare variants were defined by the same criteria as for those in the original series. Insofar as possible, genotypes from parents were used to determine whether variants were de novo or inherited.

In the COS cases, we identified and validated 27 previously unreported deletions and duplications in 23 individuals. A 2.5-MB deletion on chromosome 2q31.2-q31.3 and a 183-kb duplication on Yq11.221 were not inherited from confirmed parents and were therefore presumptively de novo; other events were either inherited or of unknown origin. A 115-kb deletion on chromosome 2p16.3 disrupting *NRXN1* was found in identical twins concordant for COS. Deletions in *NRXN1* have been reported in other patients with schizophrenia (*10*), mental retardation (*19*), and autism (*20*). In addition, two COS cases harbored the recurrent microduplication of 500-kb on chromosome 16p11.2 that was recently associated with autism and also was detected in two individuals with bipolar disorder (*21*, *22*). As found with the 16p11.2 duplication in autism (*21*, *22*), some duplications and deletions in COS patients were inherited from unaffected parents.

Among the 83 individuals with COS, 23 (28%) harbored one or more rare structural variants that deleted or duplicated genes; 4 individuals carried more than one such event (table S3). Among the 77 controls constructed from nontransmitted chromosomes of the COS cases' parents, 10 (13%) carried such variants. As for the original series, individuals with schizophrenia were significantly more likely than controls to harbor rare structural variants that altered genes (Table 1, $P = 0.03$). The proportion of controls harboring a rare deletion or duplication was higher in the COS series than in the original series. Schizophrenia of very young onset is associated with more severe family history as compared with schizophrenia of older onset (*14*), so parents of COS cases are likely to harbor more genomic lesions related to mental illness as compared with parents of persons from the general population. The use of nontransmitted parental chromosomes as controls is therefore likely to be very conservative. Deletions and duplications in COS cases and their parents were screened with different platforms from those used in the original series of cases and controls, but within each series, screening was the same for cases and controls. Definitions and sizes of CNVs are not yet identical across platforms (*13*).

Structural mutations that disrupt genes (rather than deleting or duplicating entire genes) may be especially likely to have biological consequences. Gene disruptions include premature truncation, internal deletion, gene fusion, or disruption of regulatory or promoter elements (*23*). Therefore, in our original series, we determined the breakpoints of genomic deletions and duplications in cases and controls by high-density microarray analysis and genomic sequencing (*12*). Genes disrupted by the breakpoints of each event were identified (Table 2). Individuals with schizophrenia were significantly more likely than controls to harbor rare structural variants that directly disrupted one or more genes (Table 1).

Virtually every rare structural mutation detected in our original series was different. Some deletions and duplications were multiple megabases in size; other variants altered only one or a few genes. Some events in unrelated patients altered the same genes, but with subtly different genomic breakpoints. For example, similar structural variants on chromosome 18p11 disrupted genes *LAMA1* and *PTPRM* in a case in the original series (Table 2) and in a case with COS (table S3).

Next, we assessed whether the functions of the genes disrupted in cases or controls might be related to schizophrenia. From our original series, for which disrupted genes were fully identified, we compared genes disrupted in cases versus those disrupted in controls using PANTHER (*24*) and Ingenuity Pathways Analysis (IPA) (*25*) classification systems. These programs enable one to determine whether an experimentally derived set of genes is overrepresented, as compared with all known genes (*26*), in one or more functionally defined pathways. The analyses are undirected: that is, no a priori selection is imposed on the pathways to be queried.

Genes disrupted by structural variants in our series of cases were significantly overrepresented in pathways important for brain development, including neuregulin signaling, extracellular signal–regulated kinase/mitogen-activated protein kinase (ERK/MAPK) signaling, synaptic long-term potentiation, axonal guidance signaling, integrin signaling, and glutamate receptor signaling (Table 3). Of the 24 genes disrupted by microdeletions or microduplications in cases, 11 genes—*ERBB4*, *MAGI2*, *DLG2*, *PRKCD*, *PRKAG2*, *PTK2*, *CAV1*, *GRM7*, *SLC1A3*, *PTPRM*, and *LAMA1*—were components of one or more of these pathways. Genes disrupted in controls were not overrepresented in any pathway. It is therefore important that this association was found only for cases and not for controls, that this association was not the result of the larger number of genes disrupted by case events [(*12*) and

Table 1. Individuals with schizophrenia and controls with novel structural variants (SVs) of size >100 kb. Obs, observations; OR, odds ratio; CI, confidence interval.

Subjects	*n*	Obs	Proportion with SV	*P*†	OR	95% CI	Obs	Proportion with SV	*P*†	OR	95% CI
			SVs that duplicate or delete genes					*SVs that disrupt genes*			
Cases	150	22	0.15	0.0008	3.37	1.64, 6.91	16	0.11	0.012	2.79	1.26, 6.18
Cases, onset <18 years	76	15	0.20	0.0001	4.82	2.18, 9.90	10	0.13	0.007	3.54	1.44, 8.69
Controls	268	13	0.05				11	0.04			
COS cases	83	23	0.28	0.03	2.57	1.13, 5.83					
COS controls*	77	10	0.13								

*COS controls: nontransmitted chromosomes from parents of COS cases. †Two-tailed Fisher Exact Tests for comparisons of numbers of cases versus numbers of controls with event.

table S4], and that the signaling pathways implicated were specifically involved with brain development.

Mutations in two genes, *ERBB4* and *SLC1A3*, are predicted to have particularly deleterious effects on function. For each of these genes, we experimentally characterized mutant transcripts in the affected individuals. The *ERBB4* gene, which encodes a type I transmembrane tyrosine kinase receptor for neuregulins, is disrupted by a 399-kb deletion in one case (Fig. 1A). Using 3′ rapid amplification of cDNA ends, we determined that the mutant allele produces an alternate transcript in which exons 1 to 19 are spliced to 198 base pairs (bp) of the intergenic sequence located downstream of *ERBB4* and the deleted genomic region (Fig. 1B). The alternate transcript is predicted to encode amino acids 1 to 767 of ERBB4 followed by a novel extension of 15 amino acids, which would result in a protein lacking most of the intracellular kinase domain (Fig. 1C). The predicted mutant protein is likely to function similarly to an engineered dominant-negative ERBB4 (DN-ErbB4) (Fig. 1D), which causes defects in neuronal migration (*27*) and synaptic neurotransmission (*28*, *29*).

SLC1A3, a glutamate transporter widely expressed in glial cells, regulates neurotransmitter concentrations at excitatory glutaminergic synapses (*30*). Elevated levels of *SLC1A3* message have been described in postmortem thalamus tissue of individuals with schizophrenia (*31*). The 503-kb deletion in another case disrupts both *SLC1A3* and *SKP2* and leads to a chimeric transcript expressed in patient lymphoblasts (Fig. 1E). The predicted fusion protein is likely to differ in expression and function from that of wild-type SLC1A3.

These genes and nine other genes disrupted by structural variants in cases contribute to cellular signaling networks critical to neuronal cell growth, migration, proliferation, differentiation, apoptosis, and synapse formation (*32*). ERBB4 is a receptor for neuregulin (NRG1) and interacts with MAGI2 at neuronal synapses (*33*) and with DLG2 (*34*). The NRG1-ERBB4 complex helps regulate neuronal migration and differentiation, neurotransmitter receptor expression, glial proliferation, and synaptic plasticity (*35*) and is critical to the development of glutamate networks (*28*, *29*). GRM7 [a G protein–coupled metabotropic glutamate receptor (*36*)] and SLC1A3 [a glutamate transporter (*30*)] are also involved with glutamate signaling. Glutamate is the primary excitatory neurotransmitter (*37*) in the human brain. Glutamate pathways coregulate intracellular signaling cascades, including ERK/MAPK (*38*), nitric oxide (*39*), and integrin (*40*) signaling pathways: all of which play key roles in the orchestration of neurogenesis and

Table 2. Novel SVs in genomic DNA that delete (del) or duplicate (dup) genes in schizophrenia cases and controls. Chr, chromosome; hg18, human genome assembly 18 (March 2006). Dashes indicate that no gene was disrupted by SV breakpoints.

Chr	Start (hg18)	End (hg18)	Size (bp)	Type	Duplicated or deleted genes	Genes disrupted by SV breakpoints	Age onset (years)
			Cases (n = *150*)				
1	144,943,150	146,292,286	1,349,136	del	11	*NBPF10*	25*
1	232,464,245	232,673,340	209,095	dup	3	*SLC35F3*, *TARBP1*	24
2	48,648,754	49,319,683	670,929	dup	3	*STON1-GTF2A1L*	25*
2	211,792,494	212,191,651	399,157	del	1	*ERBB4*	13
3	7,177,597	7,314,117	136,520	del	1	*GRM7*	17
3	53,056,517	53,191,698	135,181	dup	2	*PRKCD*	22
3	197,224,662	198,573,215	1,348,553	del	20	-	18
5	36,190,704	36,693,387	502,683	del	4	*SKP2*, *SLC1A3*	13
7	77,358,702	77,857,149	498,447	dup	2	*MAGI2*, *PHTF2*	18
7	100,298,156	115,966,446	15,668,290	dup	82	*SLC12A9*, *CAV1*	18
7	151,069,763	151,531,755	461,992	dup	4	*PRKAG2*, *MLL3*	15
8	142,025,432	142,393,948	368,516	dup	3	*PTK2*	12
9	2,013,316	3,118,374	1,105,058	dup	4	*SMARCA2*	>18
9	3,104,250	3,544,339	440,089	del	1	-	11
9	25,325,531	25,852,420	526,889	dup	1	-	13
11	33,260,889	33,540,102	279,213	dup	2	*HIPK3*, *C11orf41*	22
11	83,680,969	83,943,977	263,008	del	1	*DLG2*	25
14	53,485,363	53,768,132	282,769	dup	1	-	18
18	7,070,926	7,565,943	495,017	dup	2	*LAMA1*, *PTPRM*	16
19	59,045,962	59,363,706	317,744	dup	13	*TMC4*	16
22	32,048,581	32,715,286	666,705	dup	1	*LARGE*	13
Y	1	57,772,954	57,772,954	dup	entire Y	-	18
Y	57,571,543	57,723,933	152,391	dup	1	-	19
			Controls (n = *268*)				
1	99,984,742	100,204,503	219,761	del	2	*FRRS1*	
2	8,171,653	8,581,679	410,026	del	1	-	
3	79,667,848	81,408,551	1,740,703	del	1	*ROBO1*	
6	95,730,630	96,143,673	413,043	del	1	*MANEA*	
7	84,336,008	84,699,929	363,921	del	1	-	
7	111,033,890	112,334,786	1,300,896	dup	6	*FLJ31818*	
7	127,083,558	127,422,401	338,843	del	1	*SND1*	
8	8,144,616	11,755,764	3,611,148	del	20	*CTSB*	
9	12,641,413	13,216,945	575,532	dup	3	*MPDZ*	
12	24,583,650	25,248,947	665,297	dup	7	*SOX5*, *LYRM5*	
12	29,793,043	29,990,037	196,994	del	1	*TMTC1*	
16	69,395,935	69,748,827	352,892	del	2	*HYDIN*	
22	30,898,828	31,191,488	292,660	dup	5	*BPIL2*	

*Same individual.

Table 3. Pathways and processes overrepresented by genes disrupted in schizophrenia cases. No pathways were overrepresented by genes disrupted in controls.

Pathway or process	*P* value
Signal transduction*	0.012
Neuronal activities*	0.049
Nitric oxide signaling†	0.0002
Synaptic long-term potentiation†	0.0005
Glutamate receptor signaling†	0.003
ERK/MAPK signaling†	0.004
Phosphatase and tensin homolog signaling†	0.007
Neuregulin signaling†	0.008
Insulin-like growth factor 1 signaling†	0.008
Axonal guidance signaling†	0.015
Synaptic long-term depression†	0.017
G protein–coupled receptor signaling†	0.034
Integrin signaling†	0.036
Ephrin receptor signaling†	0.042
Sonic hedgehog signaling†	0.044

*Undirected PANTHER analysis (*17*). †Undirected IPA (*18*).

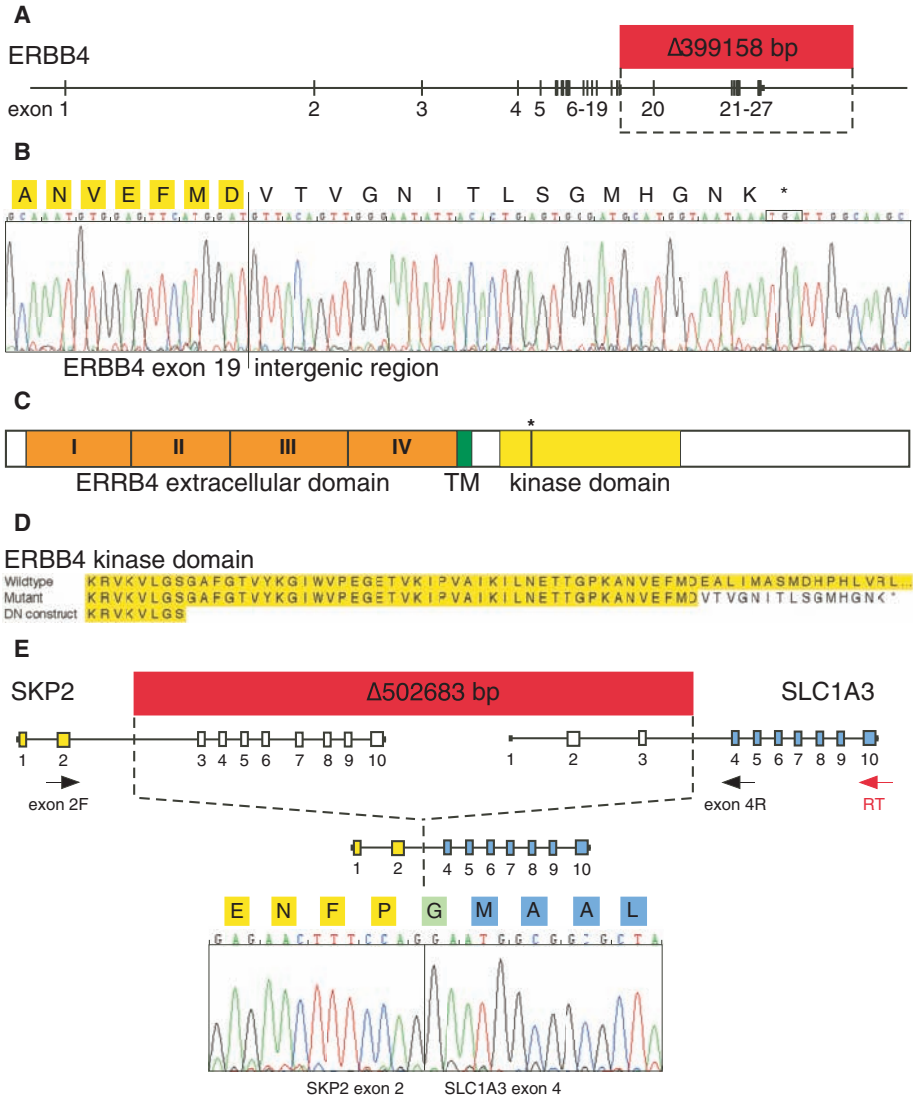

Fig. 1. Structural mutations in genomic DNA from schizophrenia patients leading to expression of truncated *ERBB4* and to expression of an in-frame fusion of *SKP2* and *SLC1A3*. (**A**) Genomic structure of the 399-kb deletion in *ERBB4* in a schizophrenia case. (**B**) cDNA sequence of the mutant *ERBB4* transcript detected in the patient's lymphoblasts, in which the sequence from exon 19 is spliced to the downstream intergenic sequence, with encoded amino acids indicated above (*47*). (**C**) Domain structure of the ERBB4 protein with the site of mutation indicated (*). TM, transmembrane domain. (**D**) Amino acid sequence of the wild-type ERBB4 kinase domain as compared with that of mutant ERBB4 and of an engineered dominant-negative ERBB4 construct (DN construct) that disrupts signaling (*27*). The predicted mutant protein has a truncated kinase domain and a novel 15 amino acid extension followed by a stop. Kinase domain amino acids are highlighted in yellow. (**E**) Chimeric *SKP2-SLC1A3* gene formed by a 503-kb deletion in a schizophrenia case. The indicated fusion transcript, including exons 1 and 2 of *SKP2* (yellow) and exons 4 to 10 of *SLC1A3* (blue), was detected in the patient's lymphoblasts. The cDNA sequence and encoded amino acids surrounding the point of fusion are shown. Arrows indicate locations of primers used to amplify the fusion region by reverse transcription polymerase chain reaction.

neural circuitry (*41*–*43*). Aberrations of NRG1-ERBB4 pathways and of glutamate neurocircuitry have been implicated previously in schizophrenia. *NRG1* is among the best-supported susceptibility genes in association studies of schizophrenia (*1*). Variants of *ERBB4* and interactions among genes in the NRG1-ERBB signaling pathway have also been associated with the illness (*44*–*46*).

In two independent cohorts, we found structural variants altering genes to be more common among individuals with schizophrenia than among controls. Our design does not prove the involvement with the illness of any specific variant or even the involvement of any specific gene. Rather, these results suggest that schizophrenia can be caused by rare mutations that disrupt genes in pathways of neuronal development and regulation. This model suggests a new approach to gene discovery for schizophrenia and likely for other psychiatric disorders. Neurodevelopmental pathways involve hundreds of genes. A severe mutation in any one of these genes may lead to a psychopathological phenotype. Furthermore, independent mutations in the same gene or genomic region, such as 16p11.2 or *NRXN1*, may lead to expression of different neurodevelopmental phenotypes, ranging from schizophrenia to autism to mental retardation (*10*, *19*–*21*) or to no clinical manifestation at all. Variable expression may arise by different genetic or epigenetic "second hits" or simply by chance.

Gene discovery efforts should focus on methods that allow detection of structural mutations genome-wide in affected individuals. Any gene harboring a deleterious structural mutation becomes a candidate gene to be screened by other methods for additional, presumably smaller, deleterious mutations in unrelated individuals. This second analytic step would entail complete analysis of variation of a limited number of candidate genes in a very large number of cases and controls. The underlying hypothesis is that a gene harboring one mutation for an illness is likely to harbor more than one mutation. Although each mutation may be individually rare, the total number of disease-causing variants in a gene relevant to the disorder may collectively explain a substantial number of cases. The long-term goals are to identify all genes with mutations leading to the illness and to develop treatment and prevention strategies tailored toward the remediation of the altered pathways.

References and Notes

1. P. J. Harrison, D. R. Weinberger, *Mol. Psychiatry* **10**, 40 (2005).
2. K. E. Lohmueller, C. L. Pearce, M. Pike, E. S. Lander, J. N. Hirschhorn, *Nat. Genet.* **33**, 177 (2003).

3. J. M. McClellan, E. Susser, M.-C. King, *Br. J. Psychiatry* **190**, 194 (2007).
4. J. A. Lee, J. R. Lupski, *Neuron* **52**, 103 (2006).
5. J. Sebat *et al.*, *Science* **316**, 445 (2007).
6. D. J. MacIntyre, D. H. R. Blackwood, D. J. Porteous, B. S. Pickard, W. J. Muir, *Mol. Psychiatry* **8**, 275 (2003).
7. J. E. Chubb, H. J. Bradshaw, D. C. Soares, D. J. Porteous, J. K. Millar, *Mol. Psychiatry* **13**, 36 (2008).
8. D. Kamnasaran, W. J. Muir, M. A. Ferguson-Smith, D. W. Cox, *J. Med. Genet.* **40**, 325 (2003).
9. H. Liu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16859 (2002).
10. G. Kirov *et al.*, *Hum. Mol. Genet.* **17**, 458 (2008).
11. J. I. Friedman *et al.*, *Mol. Psychiatry* **13**, 261 (2008).
12. Materials and methods are available as supporting material on *Science* Online.
13. A. J. Iafrate *et al.*, *Nat. Genet.* **36**, 949 (2004).
14. J. L. Rapoport, G. Inoff-Germain, *Curr. Psychiatry Rep.* **2**, 410 (2000).
15. A. Sporn *et al.*, *Mol. Psychiatry* **9**, 225 (2004).
16. J. R. Idol, A. M. Addington, R. T. Long, J. L. Rapoport, E. D. Green, *J. Autism Dev. Disord.* **38**, 668 (2008).
17. J. L. Seal *et al.*, *J. Med. Genet.* **43**, 887 (2006).
18. A. J. Sharp *et al.*, *Am. J. Hum. Genet.* **77**, 78 (2005).
19. J. M. Friedman *et al.*, *Am. J. Hum. Genet.* **79**, 500 (2006).
20. The Autism Genome Project Consortium, *Nat. Genet.* **39**, 319 (2007).
21. R. A. Kumar *et al.*, *Hum. Mol. Genet.* **17**, 628 (2007).
22. L. A. Weiss *et al.*, *N. Engl. J. Med.* **358**, 667 (2008).
23. J. R. Lupski, P. Stankiewicz, *PLoS Genet.* **1**, e49 (2005).
24. H. Mi, N. Guo, A. Kejariwal, P. D. Thomas, *Nucleic Acids Res.* **35**, D247 (2007).
25. Ingenuity Systems, www.ingenuity.com.
26. National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, www.ncbi.nlm.nih.gov.
27. C. Rio, H. I. Rieff, P. Qi, T. S. Khurana, G. Corfas, *Neuron* **19**, 39 (1997).
28. R. S. Woo *et al.*, *Neuron* **54**, 599 (2007).
29. B. Li, R. S. Woo, L. Mei, R. Malinow, *Neuron* **54**, 583 (2007).
30. P. M. Beart, R. D. O'Shea, *Br. J. Pharmacol.* **150**, 5 (2007).
31. R. E. Smith, V. Haroutunian, K. L. Davis, J. H. Meador-Woodruff, *Am. J. Psychiatry* **158**, 1393 (2001).
32. K. M. Giehl, *Prog. Exp. Tumor Res.* **39**, 1 (2007).
33. J. D. Buxbaum *et al.*, *Mol. Psychiatry* **13**, 162 (2008).
34. R. A. Garcia, K. Vasudevan, A. Buonanno, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3596 (2000).
35. R. M. Esper, M. S. Pankonin, J. A. Loeb, *Brain Res. Rev.* **51**, 161 (2006).
36. A. Makoff, C. Pilling, K. Harrington, P. Emson, *Brain Res. Mol. Brain Res.* **40**, 165 (1996).
37. M. Nedergaard, T. Takano, A. J. Hansen, *Nat. Rev. Neurosci.* **3**, 748 (2002).
38. J. Q. Wang, E. E. Fibuch, L. Mao, *J. Neurochem.* **100**, 1 (2007).
39. S. Sammut, D. J. Park, A. R. West, *J. Neurochem.* **103**, 1145 (2007).
40. J. A. Bernard-Trifilo *et al.*, *J. Neurochem.* **93**, 834 (2005).
41. J. B. Manent, A. Represa, *Neuroscientist* **13**, 268 (2007).
42. J. I. Heng, G. Moonen, L. Nguyen, *Eur. J. Neurosci.* **26**, 537 (2007).
43. C. A. Ghiani *et al.*, *Neurochem. Res.* **32**, 363 (2007).
44. N. Norton *et al.*, *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **141**, 96 (2006).
45. A. J. Law, J. E. Kleinman, D. R. Weinberger, C. S. Weickert, *Hum. Mol. Genet.* **16**, 129 (2007).
46. I. Benzel *et al.*, *Behav. Brain Funct.* **3**, 31 (2007).

47. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

48. We thank M. Wigler, E. Susser, and J. Watson for helpful discussion. This work was supported by The Forrest C. and Frances H. Lattner Foundation; a NARSAD Young Investigator Award; The NARSAD Constance and Steven Lieber Distinguished Investigator Award; The Simons Foundation; a gift of Ted and Vada Stanley; The Stanley Medical Research Foundation; the Howard Hughes Medical Institute; NIH grants HD043569, MH061464, MH061528, MH061355, NS052108, RR025014, and RR000046; NIH intramural research programs of the National Institute on Aging and the National Institute of Mental Health; and the Mental Health Division, Washington State Department of Social and Health Services.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1155174/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 to S4
References

2 January 2008; accepted 14 March 2008
Published online 27 March 2008;
10.1126/science.1155174
Include this information when citing this paper.

Metabolic Diversification—Independent Assembly of Operon-Like Gene Clusters in Different Plants

Ben Field and Anne E. Osbourn*

Operons are clusters of unrelated genes with related functions that are a feature of prokaryotic genomes. Here, we report on an operon-like gene cluster in the plant *Arabidopsis thaliana* that is required for triterpene synthesis (the thalianol pathway). The clustered genes are coexpressed, as in bacterial operons. However, despite the resemblance to a bacterial operon, this gene cluster has been assembled from plant genes by gene duplication, neofunctionalization, and genome reorganization, rather than by horizontal gene transfer from bacteria. Furthermore, recent assembly of operon-like gene clusters for triterpene synthesis has occurred independently in divergent plant lineages (*Arabidopsis* and oat). Thus, selection pressure may act during the formation of certain plant metabolic pathways to drive gene clustering.

Triterpenes protect plants against pests and diseases and are also important drugs and anticancer agents (*1*–*4*). Like sterols, these compounds are synthesized from the isoprenoid pathway by cyclization of 2,3-oxidosqualene (*1*, *3*). The *Arabidopsis* genome contains 13 predicted oxidosqualene cyclase (OSC) genes (*3*, *5*). Of these, one encodes cycloartenol synthase (CAS), which is required for sterol biosynthesis, and another encodes lanosterol synthase (LAS), which is conserved across the eudicots and whose function in plants is unknown (Fig. 1A). The 11 remaining OSCs fall into two major clades (I and II) (Fig. 1A). These OSCs produce various triterpenes when expressed in yeast. However, their function in *Arabidopsis* is unknown. The OSCs in clade I have close homologs in other eudicots; those in clade II appear to be restricted to the Brassicaceae family and show homology to a single OSC from *Brassica napus*.

Oat (*Avena* spp.), a monocot that diverged from the eudicots ~180 million years ago, produces defense-related triterpenes known as avenacins. The first committed step in avenacin synthesis is catalyzed by the OSC β-amyrin synthase (encoded by *Sad1*) (*6*). *Sad1* is hypothesized to have arisen from a duplicated monocot CAS-like gene after the separation of wheat and oat ~25 million years ago (*6*, *7*). The second step in avenacin biosynthesis is mediated by SAD2, a member of the newly described monocot-specific CYP51H subfamily of cytochrome P450 enzymes (CYP450s) (*8*). *Sad1* and *Sad2* are embedded in a gene cluster that includes genes required for acylation, glucosylation, and other steps in the pathway (*2*, *7*). The avenacin biosynthesis genes are tightly regulated and expressed only in the root epidermis, the site of accumulation of the pathway end product (*6*, *8*). The avenacin gene cluster lies within a region of the oat genome lacking synteny with rice and other cereals (*7*).

We examined the genomic regions around each of the 13 *Arabidopsis* OSC genes in the *Arabidopsis* genome to establish whether genes for triterpene synthesis might be clustered (*9*). Four OSC genes are flanked by genes predicted to encode other classes of enzymes implicated in secondary metabolism. These four OSCs all belong to clade II, which appears to have undergone accelerated evolution compared with other *Arabidopsis* OSCs (Fig. 1A). We focused on a region containing four contiguous genes predicted to encode an OSC (*At5g48010*), two CYP450s (*At5g48000* and *At5g47990*), and a BAHD family acyltransferase (ACT) (*At5g47980*) (Fig. 1B). The expression of all four genes is highly correlated (Fig. 1C) and occurs primarily in the root epidermis (fig. S1), which suggests that the genes are functionally related (*10*).

The OSC gene within this region, *At5g48010*, converts 2,3-oxidosqualene to the triterpene thalianol when expressed in yeast (*11*). However, thalianol has not been reported in plants. We detected low levels of thalianol in roots but not leaves of wild-type *Arabidopsis* (Fig. 2, C and D), consistent with the expression of *At5g48010*. Thalianol was not detectable in root extracts of a null insertion mutant of *At5g48010* (*thas1-1*) (Fig. 2E), which indicated that the *At5g48010* gene product [hereafter, named thalianol synthase (THAS)] is required for synthesis of thalianol in *Arabidopsis*

Department of Metabolic Biology, John Innes Centre, Colney Lane, Norwich NR4 7UH, UK.

*To whom correspondence should be addressed. E-mail: anne.osbourn@bbsrc.ac.uk