

Multiple variants aggregate in the neuregulin signaling pathway in a subset of schizophrenia patients

A Hatzimanolis¹, JA McGrath¹, R Wang¹, T Li², PC Wong², G Nestadt¹, PS Wolyniec¹, D Valle³, AE Pulver¹ and D Avramopoulos^{1,3}

Despite the strongly held view that schizophrenia (SZ) shows substantial genetic heterogeneity, pathway heterogeneity, as seen in cancer where different pathways are affected in similar tumors, has not been explored. We explore this possibility in a case-only study of the neuregulin signaling pathway (NSP), which has been prominently implicated in SZ and for which there is detailed knowledge on the ligand- and receptor-processing steps through β - and γ -secretase cleavage. We hypothesize that more than one damaging variants in the NSP genes might be necessary to cause disease, leading to an apparent clustering of such variants in only the few patients with affected NSP. We analyze linkage and next-generation sequencing results for the genes encoding components of the pathway, including *NRG1*, *NRG3*, *ERBB4*, β -secretase and the γ -secretase complex. We find multiple independent examples of supporting evidence for this hypothesis: (i) increased linkage scores over NSP genes, (ii) multiple positive interlocus correlations of linkage scores across families suggesting each family is linked to either many or none of the genes, (iii) aggregation of predicted damaging variants in a subset of individuals and (iv) significant phenotypic differences of the subset of patients carrying such variants. Collectively, our data strongly support the hypothesis that the NSP is affected by multiple damaging variants in a subset of phenotypically distinct patients. On the basis of this, we propose a general model of pathway heterogeneity in SZ, which, in part, may explain its phenotypic variability and genetic complexity.

Translational Psychiatry (2013) 3, e264; doi:10.1038/tp.2013.33; published online 28 May 2013

Introduction

Schizophrenia (SZ) is a common and disabling mental disorder with complex inheritance and an estimated heritability exceeding 80%.¹ Genome-wide association studies are beginning to identify common genetic variation accounting for a small fraction of the genetic risk and support the existence of a very large number of susceptibility loci.^{2–5} Many studies have demonstrated a strong selective disadvantage for SZ, the patients having on average half the number of offspring than the general population,^{6–8} yet the frequency of the disease persists high at 1%. Not surprisingly, studies also implicate *de novo* mutations in SZ,^{9–15} which suggests constant replenishment of pathogenic alleles balancing the selection process. The frequent involvement of *de novo* mutations could be due, in part, to the large number of target genes involved in neurodevelopment. The development of the nervous system is an elaborate process involving the coordinated action of genes and their products to accomplish individual tasks such as the coordinated growth, migration and synapsing of neurons to form a system capable of cognitive processing. Although the biological processes accomplishing these tasks might be robust and tolerate mildly damaging variation, it is possible that accumulation of such variation in components of the same process, either by new mutation or by co-inheritance, can lead to significant neurodevelopmental

defects and SZ. Although single deleterious mutations would be quickly cleared by selection, mildly deleterious variants that aggregate to cause disease would be resistant to selection because as their frequency becomes lower, their rate of co-occurrence is reduced exponentially and in most carriers they do not lead to a selective disadvantage.

Studies in cancer have shown that mutations in genes encoding components of several different cell proliferation pathways can produce similar tumor types.^{16,17} Like cancer, SZ shows high phenotypic heterogeneity despite the commonalities responsible for considering it a single disease entity. As many processes are involved in neurodevelopment, disruption of any one might be sufficient to cause SZ. Where in cancer a second hit often occurs as a somatic mutation,¹⁸ in the case of SZ all deleterious variants would need to be cotransmitted (which might include germline *de novo* events). For variants with relatively low allele frequencies, for example, below 10–15% each, this would be a rare event (for example, <3% for two co-occurring variants, <0.5% for three variants, and so on). In addition to their allowing persistence in the population, this could explain the small observed effect sizes, the small estimated fraction of explained heritability,² the extensive phenotypic heterogeneity of SZ and the difficulties in gene mapping. Our proposed model is illustrated in Figure 1 and we refer to it as the pathway clustering model. What is

¹Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD, USA; ²Departments of Pathology, Neurology and Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA and ³McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Correspondence: Dr D Avramopoulos, Department of Psychiatry, Johns Hopkins University School of Medicine, 733 North Broadway, Broadway Research Building Room 509, Baltimore, MD 21205, USA.

E-mail: adimitr1@jhmi.edu

Keywords: clustering; heterogeneity; linkage; neuregulin; schizophrenia; signaling pathway

Received 18 December 2012; revised 25 February 2013; accepted 15 March 2013

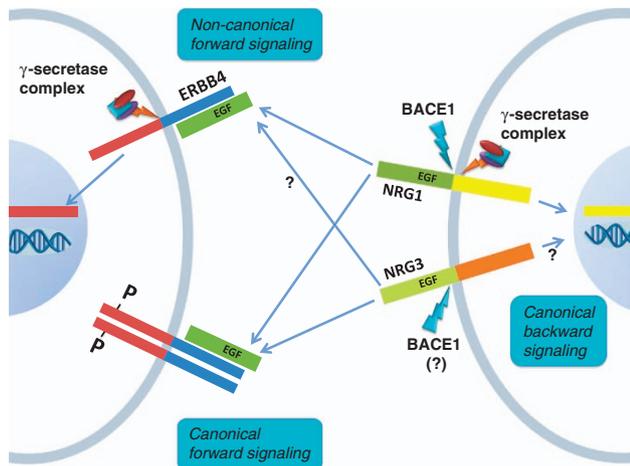


Figure 1 Schematic of neuregulin-ERBB4 signaling showing the roles of the products of the selected genes. The γ -secretase multiprotein complex is constructed from the products of the genes: *PSEN1*, *PSEN2*, *APH1A*, *APH1B*, *PSENEN* and *NCSTN*. The cleavage of NRG3 by BACE1 is hypothesized and has not been shown experimentally.

important in testing this model is to know enough about the coordinated action of genes to perform a function to define the most concise group without missing important components.

Since being associated with SZ in 2002,^{19,20} *NRG1* (encoding neuregulin 1) is among the most widely examined genes showing converging evidence of association in multiple studies.^{21–25} The evidence has been expanding to include other genes involved in neuregulin (NRG) signaling. Our group has previously reported that variation in *NRG3* (neuregulin 3), a paralog of *NRG1*, is associated with the development of delusional symptomatology in SZ patients,²⁶ a finding replicated by other investigators.^{27,28} Accumulating evidence from association studies also involves *ERBB4*, the gene encoding the main receptor for *NRG1* and *NRG3*,^{29,30} and variation in *ERBB4* has been reported to interact with variation in *NRG1* increasing susceptibility to SZ.^{31,32}

NRGs are a family of signaling molecules with extracellular epidermal growth factor domains that bind to ERBB receptors to modulate neuronal migration and differentiation.³³ Two members, NRG1 and NRG3, have been linked to SZ and they share the receptor ERBB4. Studies on NRG1 have shown that once produced it is inserted in the cell membrane, where it can either signal solely to adjacent cells or be released through cleavage by β -secretase³⁴ (Figure 1). Following binding of the NRG1 epidermal growth factor domain, ERBB4 is either autophosphorylated or transphosphorylated (canonical forward signaling) or cleaved by γ -secretase to release the C-terminal intracellular domain, which can translocate to the nucleus and regulate transcription (non-canonical forward signaling). It has also been shown that NRG1 itself can be cleaved by γ -secretase with the C-terminal intracellular domain migrating to the nucleus in canonical backward signaling.^{35–37} The components and relationships of the key players in the NRG signaling pathway (NSP) are illustrated in Figure 1. The 10 genes encoding these NSP components are the focus of our study. Evidence from mouse models supports our choice of genes and a role of β - and γ -secretase in SZ

through NRG signaling. Mice lacking *BACE1*, encoding for β -secretase, exhibit SZ-like phenotypes.³⁸ Similarly, deficiency of *APH1B*, a component of γ -secretase, also causes neurobehavioral abnormalities similar to SZ.³⁹ In addition, altered levels of specific NRG1 mRNA transcripts and protein fragments have been reported in the prefrontal cortex and hippocampus of SZ patients, providing more evidence for dysregulated NRG1 processing.^{40–42} We note that NRG3 has not yet been shown to be processed through the same mechanisms as NRG1; however, we include it in our analysis because of its close relationship to *NRG1* and its involvement in SZ.

Here, we explore our hypothesis of a pathway clustering model on the NSP, taking a narrow and stringent definition of pathway that includes only components known to be connected in a well-defined biological process. Predictions stemming from this model include that: (i) evidence of linkage (allele sharing) over the genes of the pathway will be weak but consistently elevated; (ii) families linked to one locus will also be linked to the other loci encoding components of the pathway; (iii) patients with damaging DNA variants in one gene will more often also have damaging variants in other genes of the same pathway; and (iv) the subgroup of patients with the NSP pathway affected will differ in phenotype from other patients.

Patients and methods

SZ pedigrees. We examined 123 multiplex families (that is, more than one affected family member) of European Caucasian background for which genome-wide linkage data are available and reported previously.⁴³ Details on the recruitment procedures are described in Blouin *et al.*⁴⁴ All individuals were interviewed using either a modified Diagnostic Interview Schedule⁴⁵ or the Diagnostic Interview for Genetic Studies (DIGS).⁴⁶ Diagnoses were made through a consensus procedure using the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) consensus checklist criteria. We also evaluated nine factorial phenotypes derived from a principal component phenotypic analysis of 73 indicators, which has been previously described in detail.⁴⁷ These nine factors include the often-reported delusions, hallucinations, disorganization, negative, and affective factors and novel factors such as child/adolescent sociability, scholastic performance, disability/impairment and prodromal. All nine factors demonstrate significant familiarity, with the highest scores for disability/impairment (0.61), disorganization (0.60) and scholastic performance (0.51).⁴⁷

The Johns Hopkins institutional review board has approved all recruitment methods and protocols for collection of clinical data and blood samples for this study and informed consent has been obtained from all individuals.

Linkage data analysis. Genotyping, quality control procedures and linkage analysis results have been described previously.⁴³ We genotyped the 6008 single-nucleotide polymorphisms (SNPs) in the Illumina BeadArray Linkage mapping panel version 4.0 (Illumina, San Diego, CA, USA) at the Center for Inherited Disease Research at Johns Hopkins.

We calculated the non-parametric linkage (NPL) scores across the genome for each family and for the whole set of pedigrees using the analysis software MERLIN.⁴⁸ Sample-wide NPL scores across the genome were used to extract the scores at the location of each of the 10 genes of interest and to obtain their average, which was then compared with the average scores of 100 000 random draws of 10 loci from across the genome. This method, which we have described and validated in simulated and real data, tests whether a set of loci is enriched for true linkage signals, regardless of their individual statistical significance.⁴⁹ For each gene, we identified the SNP within or closest to the gene in the linkage panel, which has a density of about one SNP every 500 kb, and the NPL score reported at that location was used for the analysis.

We used the individual family NPL scores at each of the 10 loci to examine pairwise correlations between linkage scores. These involved 123 cases (families) and 10 variables (loci). We hypothesized that if multiple hits in a pathway are involved, the families that show linkage at one locus are also more likely to show linkage at another, while the remaining families will show linkage at none. We considered NPL scores between -0.5 and $+0.5$ as uninformative and coded them as 0, otherwise positive scores were coded as 1 (evidence for linkage) and negative scores as -1 (evidence against linkage), regardless of their actual magnitude to remove biases due to family size. Owing to the lack of normal distribution, Kendall's tau (τ) coefficient was used to calculate linkage correlations between loci. The same was used to calculate correlations between the number of variants at one gene and the number of variants at the remaining pathway genes. These *P*-values are one-tailed since only positive correlations are consistent with the model and were validated by 1 000 000 permutations of the variant carriers for each gene (see Supplementary Material). To assess correlations between quantitative factor phenotypes and the load of variants in the NSP, we used a generalized linear model that included family ID as a categorical factor. This was preferred because we have shown that these factors show significant familial aggregation.⁴⁷ These tests were two-tailed as there is no *a priori* expected direction. All statistical analysis was performed using the software Statistica (StatSoft, Tulsa, OK, USA).

Sample pooling and exon sequencing. We sequenced pairs of distant affected relatives, so that in addition to assessing clustering of variants in individuals, we could assess whether the same pathway is perturbed across the extended family. On the basis of the availability of affected distant relatives and regardless of linkage to the loci, we selected 24 pairs of relatives (48 SZ cases), the closest relationship being avuncular (25% expected genome sharing). The expected identity by descent due to relatedness was 0.25 for 12 pairs, 0.125 for 9 pairs, 0.0625 for 2 pairs and 0.03125 for 1 pair. We used next-generation technology to sequence all 120 protein coding exons (including alternative exons) of *NRG1* and *NRG3*, the NRG receptor gene *ERBB4*, the *BACE1* gene, which encodes β -secretase enzyme, and the six genes that encode the components of the γ -secretase multiprotein complex (*PSEN1*, *PSEN2*, *PSENE1*, *NCSTN*, *APH1A* and *APH1B*), totaling 38.9 kb of genomic sequence. We sequenced amplified DNA from pools of four subjects (12 pools in total) following the procedures reported by Niranjana *et al.*⁵⁰ with minor modifications. We aligned the reads to the reference human genome sequence (hg19 build) using the Bowtie alignment program,⁵¹ and subsequently processed, sorted and indexed using SAMtools.⁵² Variant identification was performed for each pool using the Integrated Genomics Viewer software (IGV 2.1, <http://www.broadinstitute.org/igv/>). We validated all identified coding variants and de-convoluted the pools by Sanger sequencing (detailed procedures in Supplementary Material).

Results

Linkage over the 10 loci comprising the NSP. We have previously demonstrated that testing the average linkage score across multiple loci and comparing it to the average of randomly selected loci is a powerful approach to show enrichment in true loci.⁴⁹ According to the pathway clustering model, we expect such enrichment across the 10 NSP genes. The individual NPL scores from our 123 multiplex pedigrees are as follows: *BACE1*, 1.92; *APH1A*, 1.70; *PSENE1*, 1.21; *NCSTN*, 1.17; *APH1B*, 0.96; *NRG1*, 0.92; *NRG3*, 0.12; *ERBB4*, -0.25 ; *PSEN2*, -0.44 ; and *PSEN1*,

Table 1 Pairwise correlations between linkage signals across 123 multiplex schizophrenia families

GENE	NCSTN	PSEN2	APH1A	PSEN1	APH1B	PSENE1	BACE1	NRG1	NRG3	ERBB4
<i>NCSTN</i>		<u>0.17</u>		<u>0.12</u>	<u>0.15</u>	0.05	<u>0.17</u>	-0.06	0.04	0.08
<i>PSEN2</i>	0.005		0.12	0.04	0.00	0.12	0.07	-0.01	<u>-0.17</u>	0.06
<i>APH1A</i>		0.046		0.05	0.07	-0.06	<u>0.13</u>	0.00	0.08	0.07
<i>PSEN1</i>	0.046				0.07	0.03	-0.09	0.05	-0.01	0.02
<i>APH1B</i>	0.014					0.02	-0.06	-0.03	0.01	-0.01
<i>PSENE1</i>							0.07	-0.05	<u>-0.13</u>	-0.01
<i>BACE1</i>	0.006		0.028					-0.03	0.03	0.01
<i>NRG1</i>									0.03	0.03
<i>NRG3</i>		0.005				0.036			-0.05	0.00
<i>ERBB4</i>										

Kendall's τ coefficient is shown above the diagonal, with nominally significant correlations underlined and those that are positive in bold. Nominally significant *P*-values are below the diagonal and if the correlation is positive they are in bold. The *P*-value for the linked loci NCSTN-APH1A has been removed.

–0.59. The overall average is 0.67. To assess the significance of this combined linkage, we performed 100 000 draws of 10 loci at random from the same genome scan providing an empirical $P=0.067$, which, although not formally reaching significance, suggests that many of the 10 NSP genes may carry SZ risk alleles.

Correlation of linkage scores across SZ families.

According to our pathway clustering model, the subset of families with disrupted NSP function will show linkage to many NSP loci, while the other families to none, generating correlations of linkage scores between NSP loci. All pairwise correlations of linkage scores between the 10 NSP loci in our data are shown in Table 1. Two genes, *NCSTN* and *APH1A*, are located <50cM apart on the same chromosome; therefore; correlation of linkage is expected on this basis and should be disregarded (the P -value has been removed from Table 1). Of the 44 remaining comparisons, 28 showed positive and 16 negative correlations compared to an expected 22 of each (observed vs expected $\chi^2 P=0.07$). Two nominally significant negative correlations were observed, both for *NRG3* (not different from the 1.1 negative correlation expected by chance in 44 tests, underlined in

Table 1). In contrast, six positive correlations were statistically significant (Kendall's rank correlation $P<0.05$, bold and underlined in Table 1). The number of positive nominally significant results is almost sixfold higher than expected by chance (6 observed vs 1.1 expected; $\chi^2 P=2.2 \times 10^{-6}$), consistent with our model. As it could be argued that the positional correlation between *NCSTN* and *APH1A* makes the observations non-independent, we repeated the analysis removing all comparisons involving *APH1A* and this enrichment remained significant ($P=7.2 \times 10^{-4}$). This independent evidence from two different approaches prompted us to search for the responsible variants by sequencing all coding exons of the 10 NSP genes.

Clustering of damaging coding variants in the NSP.

We applied next-generation sequencing on 24 pairs of distantly related affected subjects in 12 tagged pools of 4. We obtained on average 12 356 105 100-bp single-end reads from each indexed library and more than 75% of the reads mapped successfully to the hg19 human reference genome using Bowtie default parameters. The mean coverage for coding variants across all 12 libraries was 21 613 (median 21 504) reads, which resulted in an average read depth of 2702 reads per allele. To complete our screen, we sequenced all exons with $<20 \times$ coverage per individual allele (2.6% of all exons) by Sanger sequencing in the respective pools.

All identified exonic variants are shown in Supplementary Table 2. We detected 13 non-synonymous missense variants in 15 of the 24 pairs examined (62.5%). We validated all variants and de-convoluted the pools by conventional Sanger sequencing. Twelve variants were present in dbSNP135 and one, located in *APH1B* gene, was novel (p.Val199Leu). Characterization of the functional impact of coding variants using PolyPhen-2 and SIFT algorithms^{53,54} predicted that 7 out of the 13 are potentially deleterious to protein function. All predicted damaging variants had minor allele frequencies below 0.12 except rs3924999 in *NRG1*, which had a minor allele frequency of 0.48. On the basis of our hypothesis as outlined in the introduction, we only considered for further analysis variants that had a minor allele frequency below 0.12 and were predicted to be damaging by at least one algorithm. These are shown in Supplementary Table 2.

Table 2 Correlations between the number of damaging variants in each gene, with the number of variants in the remaining NSP genes

	Individuals (N = 48)		Families (N = 24)	
	Kendall τ	P-value (one-tailed)	Kendall τ	P-value (one-tailed)
<i>NRG1</i> vs other	0.23	0.010	0.31	0.016
<i>APH1B</i> vs other	0.28	0.002	0.33	0.011
<i>NRG3</i> vs other	–0.17	NA	–0.24	NA
<i>Excluding NRG3 variants</i>				
<i>NRG1</i> vs other	0.34	3.6×10^{-4}	0.53	1.5×10^{-4}
<i>APH1B</i> vs other	0.50	2.8×10^{-7}	0.67	2.5×10^{-6}

Abbreviations: NA, not applicable; NRG, neuregulin; NSP, neuregulin signaling pathway; SZ, schizophrenia.

Rare variant correlations for each gene with the remaining NSP genes based on 48 SZ cases. To account for familial genotype correlations, results are also shown for the average variant load of each pair (24 observations). The lower two lines show the results after excluding *NRG3* from the NSP (see text).

Table 3 Differences in symptom dimensions between those SZ cases that carry NSP damaging variants and those that do not

Factor	Number of NSP variants		Excluding <i>NRG3</i>		<i>NRG3</i> alone	
	Direction	P-value	Direction	P-value	Direction	P-value
Negative symptoms		0.586		0.376		0.623
Affective symptoms		0.982		0.944		0.859
Disorganization		0.420		0.490		0.677
Prodromal signs		0.706		0.874		0.613
Delusions		0.231		0.521		0.183
Scholastic functioning		0.791		0.444		0.386
Child/adolescent sociability	(–)	0.034	(–)	0.015	(–)	0.998
Impairment/disability	(–)	0.018	(–)	0.004	(+)	0.850
Hallucinations	(+)	0.141	(+)	0.028	(–)	0.380

Abbreviations: NRG, neuregulin; NSP, neuregulin signaling pathway; SZ, schizophrenia.

The effect of NSP damaging variants on the phenotype is measured through correlations between NSP variant load and factorial dimension scores (see text). The direction of the effect is only shown for significant P -values which are shown in bold.

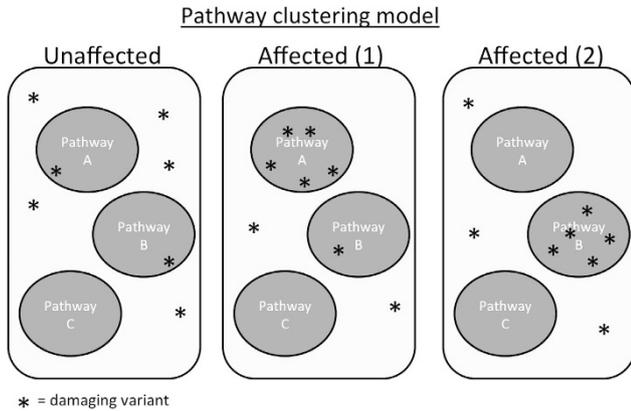


Figure 2 Pathway clustering model. Schematic representation of the proposed pathway clustering model for schizophrenia demonstrating how clustering of multiple protein damaging variants within different well-defined molecular pathways could represent an unrecognized factor of SZ risk and phenotypic heterogeneity.

Our model predicts that potentially deleterious variants should cluster in certain SZ individuals/pedigrees, those where the disease is caused by disruption of the NSP. The question therefore is whether a SZ patient who has damaging variants in one NSP gene is more likely to also have damaging variants in the other NSP genes. We tested by Kendall's rank correlation analysis for correlations between the counts of variants in each gene compared with the count of variants in the remaining genes. This is a total of three tests, one for each of *NRG1*, *APH1B* and *NRG3*; *APH1A* was not tested because only two individuals carried the identified mutation, but was included when testing other genes. We found significant correlations for *NRG1* and *APH1B* ($\tau=0.23$, $P=0.01$, and $\tau=0.28$, $P=0.002$, respectively, one-tailed). *NRG3* showed a nonsignificant negative correlation. To account for within-family genotype correlations that might inflate the statistic (the 48 observations are not truly independent), we also performed the analysis on the average variant burden of the two relatives and the family as a unit. Despite reducing the observations to 24, we obtained similar strong results (Table 2).

The lack of positive correlations for *NRG3* was of interest as the link of this gene with β - and γ -secretase processing has not been shown. It is also interesting that *NRG3* did not show positive linkage correlations with other genes, whereas when we tried removing it from the set of genes tested for their NPL average, the permutation P -value improved ($P=0.053$). Although *NRG3* is a paralog of *NRG1* and may share some pathway components, it may well function in a different or parallel pathway. For this reason, we performed the same analysis excluding *NRG3*. We observed marked improvement and high statistical significance of the correlations of variation in *APH1B* and *NRG1* with the remaining genes ($\tau=0.67$, $P=2.5 \times 10^{-6}$, permutation $P=1.7 \times 10^{-4}$; $\tau=0.53$, $P=1.5 \times 10^{-4}$, permutation $P=3.4 \times 10^{-3}$). This result, which remains significant after Bonferroni-based correction for removing each of the genes and re-testing, strongly supports our hypothesis and has potential implications for the possible relationship of *NRG3* to the NSP and its processing compared to *NRG1*. It is interesting to observe that while the

carriers of the *NRG3* variant do not have more variants in other NSP genes, in four of the six families both affected members carry the damaging variant, despite the expected identity by descent sharing being 25% or less.

Phenotypic correlates of NSP damaging variants. The pathway clustering model predicts that individuals with different affected pathways might differ in phenotype. The sequencing results above allow us to identify individuals whose SZ is likely due to variants in the NSP. The same individuals have been studied in a phenotype factor analysis that we have performed and published,⁴⁷ each having available scores for nine quantitative factors reflecting distinct clinical dimensions. We tested whether the NSP mutational load correlates with the each of the nine factor scores, the scores that are calculated for each individual based on the presence or absence of the 73 phenotypic indicators and their loadings on each factor. Since we have shown that the factors show familial aggregation,⁴⁷ for these factors we used a generalized linear model that included family ID as a categorical predictor. Owing to the sequencing results above questioning the inclusion of *NRG3* in the gene set, we performed this analysis with and without *NRG3*. We observed significant correlations with two of the nine factors that improved and increased to three when excluding *NRG3* variants (Table 3). The strongest result was for the impairment/disability factor ($P=0.004$), with lower disability in carriers of NSP damaging variants compared to patients with other pathways presumably affected. Of note, the presence of NSP damaging variants also predicted an elevated hallucination score when *NRG3* variants were not included in the analysis ($P=0.028$).

Discussion

SZ is a complex multifactorial mental disorder likely caused by the synergistic effects of genetic and environmental influences.^{2,55} As sample sizes have grown, genetic studies are detecting common polymorphic variants robustly associated with SZ in diverse ethnic populations, signifying that SZ is a highly polygenic disorder with multiple genetic loci involved.^{2,4} In line with this notion, a recent hypothesis suggests that hundreds or even thousands of loci act together in the context of inter-related molecular pathways to confer liability to SZ.⁵⁶ Here, we further propose that in individual cases SZ may be the result of variants accumulating and perturbing the function of only one or a few of the many critical pathways. The specific pathway(s) may differ between patients, but might be similar within families. This model combines pathway heterogeneity and a requirement for multiple hits to the pathway. Its main difference from models commonly used in sequencing studies for rare variants is the prediction of heterogeneity among patients, which increases power and can be explored in a patient-only design. It further proposes that in addition to the number of damaging genetic variants, their distribution in pathway gene-sets is crucial, which could explain the difficulties of association and sequencing studies in identifying causative genetic loci. Directly testing this model will soon become possible with extensive exome and genome sequencing data combined with detailed and reliable delineation of

the relevant pathways. In our study, we focus on one pathway well connected to SZ whose core components are well characterized and we consistently find support for the model.

There are two important aspects of our results, one regarding the NSP and another regarding the possibility that this model can be generalized. As others have previously suggested through association studies,^{25,28,57} we now show through different approaches that multiple members of the NSP are involved in SZ. The increased linkage scores across the NSP gene-set, the correlations between linkage scores, the aggregation of potentially deleterious variants in a subset of patients and families and the phenotypic differences of these individuals strongly support the idea that multiple hits in the NSP increase the risk of SZ in a subset of individuals and families. We point to a few specific variants that should be the subject of functional follow-up including their possible biological interactions. There is little doubt that these variants are only a subset and that others, including non-coding or synonymous functional variants, could be involved. Further, our list of NSP genes may be incomplete and it appears from our results that *NRG3* should be removed. Having a reliable model will help identify such variants and new gene-set members in larger samples, having a significant impact on our knowledge of the role of the NSP in SZ etiology. It should be noted that our assumption of a linear relationship between NSP variant load and phenotype is admittedly an oversimplification, only applied for lack of enough data to explore the true underlying model. In our study 6 out of 24 families (25%) carry variants in multiple NSP genes and we find that carrying such variants leads to lower scores in the 'impairment/disability' phenotypic factor (higher functioning) and higher scores in the 'hallucinations' factor (more positive symptoms). This is of particular interest in view of results exploring *NRG3* and showing an association with a subgroup of subjects with SZ who experience florid delusions and have relatively spared cognition.²⁷ Although our factor scores are only a statistical construct and the factor names simply derived from highly loading phenotypic items, there is a clear parallel emerging suggesting the NSP as a pathway associated with positive symptoms, relatively sparing cognition and retaining functional capacity.

The second aspect of these data is that they are consistent with a broad model of pathway heterogeneity and damaging variant clustering for SZ illustrated in Figure 2. This pathway clustering model can be exploited for the analysis of forthcoming large-scale sequencing data on SZ cases and controls. Although we have not shown that other pathways behave in the same way, the NSP provides a paradigm. In fact, a likely scenario might involve more than one biological process per pathway and more than one pathway per affected individual, in a nested clustering model that future large patient sequencing data sets will make possible to explore. Today few other pathways combine the connections to SZ and the biological knowledge available for NRG. Possible candidates include *RELN*,⁵⁸ *DISC1*-interacting proteins⁵⁹ and pathways emerging from large collaborative genome-wide association studies.⁶⁰ Statistical models based on the pathway heterogeneity hypothesis applied on forthcoming larger data sets with increased statistical power can test networks of genes revealed from proteomics,⁶¹ gene expression

experiments and gene ontology databases. Exploratory analyses could help refine these networks, greatly improving our knowledge of the disease.

While our results are positive and consistent, it is important to keep in mind the limitations of each approach. Linkage information is well known to be inherently noisy, especially when the number of families is relatively small. While our test across loci approaches significance, the individual NPL scores should be considered with caution. The same is true for positive correlations between linkage scores, which are found in highly significant excess but are hard to assess individually. Further, it cannot be assumed that we detected all deleterious variants or all the variants we analyzed are in fact deleterious. Some of our predicted functional variants could be mischaracterized and others missed, together with all functional variants in non-coding regions. This could explain why some genes in which we found no variants show strong linkage scores and correlations. Finally, as we noted it is more than likely that the conservatively selected set of NSP genes is incomplete or inaccurate. For example, *NRG3* perhaps should be examined independently from *NRG1*.

Overall, in this work we explore the NSP and find that it shows aggregation of potentially damaging genetic variation in a subset of SZ patients and families. This result leads us to propose a model of pathway clustering, where multiple hits in one of many possible pathways explains SZ in a subset of cases. Exploring this possibility could lead to a much better understanding of disease risk and genetic heterogeneity, and also provide a framework for more successful analytical approaches.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements. We thank the patients and controls for their participation. The collection of patients and controls was funded by NIMH Grants R01MH057314 and R01MH068406 (Dr Pulver). Dr Avramopoulos is supported by NIMH Grants R01MH092515 and R01MH085018. Dr Valle and Avramopoulos are supported by a grant from the Johns Hopkins Brain Science Institute.

- Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ *et al*. Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch Gen Psychiatry* 1999; **56**: 162–168.
- Lee SH, DeCandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME *et al*. Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ), International Schizophrenia Consortium (ISC). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 2012; **44**: 247–250.
- Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet* 2012; **13**: 537–551.
- Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF *et al*. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
- Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011; **43**: 969–976.
- Nimgaonkar VL. Reduced fertility in schizophrenia: here to stay? *Acta Psychiatr Scand* 1998; **98**: 348–353.
- Mansour H, Kandil K, Wood J, Fathi W, Elassy M, Ibrahim I *et al*. Reduced fertility and fecundity among patients with bipolar I disorder and schizophrenia in Egypt. *Psychiatry Invest* 2011; **8**: 214–220.
- Bundy H, Stahl D, MacCabe JH. A systematic review and meta-analysis of the fertility of patients with schizophrenia and their unaffected relatives. *Acta Psychiatr Scand* 2011; **123**: 98–106.

9. Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM *et al*. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**: 539–543.
10. Veltman JA, Brunner HG. *De novo* mutations in human genetic disease. *Nat Rev Genet* 2012; **13**: 565–575.
11. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S *et al*. Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. *Nat Genet* 2011; **43**: 864–868.
12. Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S *et al*. High frequencies of *de novo* CNVs in bipolar disorder and schizophrenia. *Neuron* 2011; **72**: 951–963.
13. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L *et al*. Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nat Genet* 2011; **43**: 860–863.
14. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat Genet* 2008; **40**: 880–885.
15. Rees E, Kirov G, O'Donovan MC, Owen MJ. *De novo* mutation in schizophrenia. *Schizophr Bull* 2012; **38**: 377–381.
16. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012; **490**: 61–70.
17. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P *et al*. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 2008; **321**: 1801–1806.
18. Knudson AG. Two genetic hits (more or less) to cancer. *Nat Rev Cancer* 2001; **1**: 157–162.
19. Stefansson H, Sarginson J, Kong A, Yates P, Steinthorsdottir V, Gudfinnsson E *et al*. Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. *Am J Hum Genet* 2003; **72**: 83–87.
20. Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S *et al*. Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* 2002; **71**: 877–892.
21. Alaerts M, Ceulemans S, Forero D, Moens LN, De Zutter S, Heyman L *et al*. Support for NRG1 as a susceptibility factor for schizophrenia in a northern swedish isolated population. *Arch Gen Psychiatry* 2009; **66**: 828–837.
22. Nicodemus KK, Law AJ, Luna A, Vakkalanka R, Straub RE, Kleinman JE *et al*. A 5' promoter region SNP in NRG1 is associated with schizophrenia risk and type III isoform expression. *Mol Psychiatry* 2009; **14**: 741–743.
23. Georgieva L, Dimitrova A, Ivanov D, Nikolov I, Williams NM, Grozeva D *et al*. Support for neuregulin 1 as a susceptibility gene for bipolar disorder and schizophrenia. *Biol Psychiatry* 2008; **64**: 419–427.
24. Harrison PJ, Law AJ. Neuregulin 1 and schizophrenia: genetics, gene expression, and neurobiology. *Biol Psychiatry* 2006; **60**: 132–140.
25. Li D, Collier DA, He L. Meta-analysis shows strong positive association of the neuregulin 1 (NRG1) gene with schizophrenia. *Hum Mol Genet* 2006; **15**: 1995–2002.
26. Chen PL, Avramopoulos D, Lasseter VK, McGrath JA, Fallin MD, Liang KY *et al*. Fine mapping on chromosome 10q22–q23 implicates neuregulin 3 in schizophrenia. *Am J Hum Genet* 2009; **84**: 21–34.
27. Morar B, Dragovic M, Waters FA, Chandler D, Kalaydjieva L, Jablensky A. Neuregulin 3 (NRG3) as a susceptibility gene in a schizophrenia subtype with florid delusions and relatively spared cognition. *Mol Psychiatry* 2011; **16**: 860–866.
28. Kao WT, Wang Y, Kleinman JE, Lipska BK, Hyde TM, Weinberger DR *et al*. Common genetic variation in neuregulin 3 (NRG3) influences risk for schizophrenia and impacts NRG3 expression in human brain. *Proc Natl Acad Sci USA* 2010; **107**: 15619–15624.
29. Silberberg G, Darvasi A, Pinkas-Kramarski R, Navon R. The involvement of ErbB4 with schizophrenia: association and expression studies. *Am J Med Genet B* 2006; **141B**: 142–148.
30. Law AJ, Kleinman JE, Weinberger DR, Weickert CS. Disease-associated intronic variants in the *ErbB4* gene are related to altered ErbB4 splice-variant expression in the brain in schizophrenia. *Hum Mol Genet* 2007; **16**: 129–141.
31. Nicodemus KK, Luna A, Vakkalanka R, Goldberg T, Egan M, Straub RE *et al*. Further evidence for association between ErbB4 and schizophrenia and influence on cognitive intermediate phenotypes in healthy controls. *Mol Psychiatry* 2006; **11**: 1062–1065.
32. Norton N, Moskvina V, Morris DW, Bray NJ, Zammit S, Williams NM *et al*. Evidence that interaction between neuregulin 1 and its receptor erbB4 increases susceptibility to schizophrenia. *Am J Med Genet B* 2006; **141B**: 96–101.
33. Falls DL. Neuregulins: functions, forms, and signaling strategies. *Exp Cell Res* 2003; **284**: 14–30.
34. Fleck D, Garratt AN, Haass C, Willem M. BACE1 dependent neuregulin processing: review. *Curr Alzheimer Res* 2012; **9**: 178–183.
35. Sardi SP, Murtie J, Koirala S, Patten BA, Corfas G. Presenilin-dependent ErbB4 nuclear signaling regulates the timing of astrogenesis in the developing brain. *Cell* 2006; **127**: 185–197.
36. Mei L, Xiong WC. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. *Nat Rev Neurosci* 2008; **9**: 437–452.
37. Bao J, Wolpowitz D, Role LW, Talmage DA. Back signaling by the nrg-1 intracellular domain. *J Cell Biol* 2003; **161**: 1133–1141.
38. Savonenko AV, Melnikova T, Laird FM, Stewart KA, Price DL, Wong PC. Alteration of BACE1-dependent NRG1/ErbB4 signaling and schizophrenia-like phenotypes in BACE1-null mice. *Proc Natl Acad Sci USA* 2008; **105**: 5585–5590.
39. Dejaegere T, Serneels L, Schafer MK, Van Biervliet J, Horre K, Depboylu C *et al*. Deficiency of Aph1B/C-gamma-secretase disturbs Nrg1 cleavage and sensorimotor gating that can be reversed with antipsychotic treatment. *Proc Natl Acad Sci USA* 2008; **105**: 9775–9780.
40. Chong VZ, Thompson M, Beltaifa S, Webster MJ, Law AJ, Weickert CS. Elevated neuregulin-1 and ErbB4 protein in the prefrontal cortex of schizophrenic patients. *Schizophr Res* 2008; **100**: 270–280.
41. Marballi K, Cruz D, Thompson P, Walss-Bass C. Differential neuregulin 1 cleavage in the prefrontal cortex and hippocampus in schizophrenia and bipolar disorder: Ppreliminary findings. *PLoS One* 2012; **7**: e36431.
42. Law AJ, Lipska BK, Weickert CS, Hyde TM, Straub RE, Hashimoto R *et al*. Neuregulin 1 transcripts are differentially expressed in schizophrenia and regulated by 5' SNPs associated with the disease. *Proc Natl Acad Sci USA* 2006; **103**: 6747–6752.
43. Holmans PA, Riley B, Pulver AE, Owen MJ, Wildenauer DB, Gejman PV *et al*. Genomewide linkage scan of schizophrenia in a large multicenter pedigree sample using single nucleotide polymorphisms. *Mol Psychiatry* 2009; **14**: 786–795.
44. Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G *et al*. Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet* 1998; **20**: 70–73.
45. Pulver AE, Wolyniec PS, Wagner MG, Moorman CC, McGrath JA. An epidemiologic investigation of alcohol-dependent schizophrenics. *Acta Psychiatr Scand* 1989; **79**: 603–612.
46. Numberger JI Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J *et al*. Diagnostic interview for genetic studies. rationale, unique features, and training. NIMH genetics initiative. *Arch Gen Psychiatry* 1994; **51**: 849–859; discussion 863–864.
47. McGrath J, Avramopoulos D, Lasseter VK, Wolyniec P, Fallin MD, Liang KY *et al*. Familiality of novel factorial dimensions of schizophrenia. *Arch Gen Psychiatry* 2009; **66**: 591–600.
48. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. MERLIN—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
49. Avramopoulos D, Zandi P, Gherman A, Fallin MD, Bassett SS. Testing groups of genomic locations for enrichment in disease loci using linkage scan data: a method for hypothesis testing. *Hum Genom* 2006; **2**: 345–352.
50. Niranjana TS, Adamczyk A, Bravo HC, Taub MA, Wheelan SJ, Irizarry R *et al*. Effective detection of rare variants in pooled DNA samples using cross-pool tailcure analysis. *Genome Biol* 2011; **12**: R93.
51. Langmead B. Aligning short sequencing reads with bowtie. *Curr Protoc Bioinform* 2010; Chapter 11, Unit 11.7.
52. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 2009; **25**: 1754–1760.
53. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P *et al*. A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.
54. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012; **40**: W452–W457.
55. van Os J, Kenis G, Rutten BP. The environment and schizophrenia. *Nature* 2010; **468**: 203–212.
56. Sullivan PF. Puzzling over schizophrenia: schizophrenia as a pathway disease. *Nat Med* 2012; **18**: 210–211.
57. Munafo MR, Thiselton DL, Clark TG, Flint J. Association of the NRG1 gene and schizophrenia: a meta-analysis. *Mol Psychiatry* 2006; **11**: 539–546.
58. Liu Y, Chen PL, McGrath J, Wolyniec P, Fallin D, Nestadt G *et al*. Replication of an association of a common variant in the Reelin gene (RELN) with schizophrenia in Ashkenazi Jewish women. *Psychiatr Genet* 2010; **20**: 184–186.
59. Kang E, Burdick KE, Kim JY, Duan X, Guo JU, Sailor KA *et al*. Interaction between FEZ1 and DISC1 in regulation of neuronal development and risk for schizophrenia. *Neuron* 2011; **72**: 559–571.
60. Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM *et al*. The International Schizophrenia Consortium. Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Mol Psychiatry* 2012; **17**: 996–1006.
61. Martins-de-Souza D. Proteomics tackling schizophrenia as a pathway disorder. *Schizophr Bull* 2012; **38**: 1107–1108.



Translational Psychiatry is an open-access journal published by **Nature Publishing Group**. This work is licensed under a **Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License**. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary Information accompanies the paper on the Translational Psychiatry website (<http://www.nature.com/tp>)