Biological Validation of Increased Schizophrenia Risk With NRG1, ERBB4, and AKT1 Epistasis via Functional Neuroimaging in Healthy Controls

Kristin K. Nicodemus, PhD, MPH; Amanda J. Law, PhD; Eugenia Radulescu, MD, PhD; Augustin Luna, BS; Bhaskar Kolachana, PhD; Radhakrishna Vakkalanka, PhD; Dan Rujescu, PhD; Ina Giegling, PhD; Richard E. Straub, PhD; Kate McGee, PhD; Bert Gold, PhD; Michael Dean, PhD; Pierandrea Muglia, MD; Joseph H. Callicott, MD; Hao-Yang Tan, MD; Daniel R. Weinberger, MD

Context: NRG1 is a schizophrenia candidate gene and plays an important role in brain development and neural function. Schizophrenia is a complex disorder, with etiology likely due to epistasis.

Objective: To examine epistasis between NRG1 and selected N-methyl-d-aspartate–glutamate pathway partners implicated in its effects, including ERBB4, AKT1, DLG4, NOS1, and NOS1AP.

Design: Schizophrenia case-control sample analyzed using machine learning algorithms and logistic regression with follow-up using neuroimaging on an independent sample of healthy controls.

Participants: A referred sample of schizophrenic patients (n=296) meeting DSM-IV criteria for schizophrenia spectrum disorder and a volunteer sample of controls for case-control comparison (n=365) and a separate volunteer sample of controls for neuroimaging (n=172).

Main Outcome Measures: Epistatic association between single-nucleotide polymorphisms (SNPs) and case-control status; epistatic association between SNPs and the blood oxygen level–dependent physiological response during working memory measured by functional magnetic resonance imaging.

Results: We observed interaction between NRG1 5’ and 3’ SNPs rs4360751 and rs3802160 (likelihood ratio test P=.00020) and schizophrenia, which was validated using functional magnetic resonance imaging of working memory in healthy controls; carriers of risk-associated genotypes showed inefficient processing in the dorsolateral prefrontal cortex (P=.015, familywise error corrected). We observed epistasis between NRG1 (rs10503929; Thr286/289/294Met) and its receptor ERBB4 (rs1026882; likelihood ratio test P=.035); a 3-way interaction with these 2 SNPs and AKT1 (rs2494734) was also observed (odds ratio, 27.13; 95% confidence interval, 3.30-223.03; likelihood ratio test P=.042). These same 2- and 3-way interactions were further biologically validated via functional magnetic resonance imaging: healthy individuals carrying risk genotypes for NRG1 and ERBB4, or these 2 together with AKT1, were disproportionately less efficient in dorsolateral prefrontal cortex processing. Lower-level interactions were not observed between NRG1/ERBB4 and AKT1 in association or neuroimaging, consistent with biological evidence that NRG1 × ERBB4 interaction modulates downstream AKT1 signaling.

Conclusion: Our data suggest complex epistatic effects implicating an NRG1 molecular pathway in cognitive brain function and the pathogenesis of schizophrenia.

Arch Gen Psychiatry. 2010;67(10):991-1001

N EUREGULIN 1 (NRG1) is considered a schizophrenia candidate gene because of its chromosomal position within a linkage peak, direct evidence supporting genetic association with various polymorphisms,1,12 and its biological role in brain development and neural function mediated partially via the N-methyl-d-aspartate (NMDA)–glutamate pathway. Association has been confirmed at both the 5’ and 3’ ends,12 although null studies have appeared.13-20 One meta-analysis24 of the original risk-associated haplotype first identified in an Icelandic sample (HAPICE) revealed a small effect size (odds ratio [OR], 1.22) but strong statistical association (P value=8.0 × 10−10), and another showed a significant, but weaker, association (P value=.036).22 However, these meta-analyses do not consider recently published genome-wide association study data that have provided less robust evidence. Although schizophrenia is thought to be a polygenic disorder, few studies of NRG1 epistasis have been conducted. There have been reports of interaction between HAPICE and its receptor ERBB4.23-25 However, none of these interactions were independently validated and recent data suggest that epistasis is likely the product of higher-order (ie, >2 single-nucleotide polymorphisms [SNPs]) interactions that these studies were not designed to detect.26

Author Affiliations are listed at the end of this article.
Machine learning algorithms (MLAs), developed for high-dimensional data, may be a viable approach to detecting higher-order epistasis that avoids some of the obstacles of more traditional statistical approaches. We applied 3 algorithms27-31 based on classification trees, which simultaneously model both main effects and interactions, to an analysis of selected genes in the NRG1-NMDA network and risk for schizophrenia. An example of the creation of a forest of classification trees is shown in Figure 1. Essentially, with a case-control outcome, trees use the predictor variables (herein SNPs) to partition the data, creating more homogeneous subsets containing a majority of cases or controls, or the algorithm searches the space of all models to try to find the best-fitting tree. Algorithms are described in more detail in the “Methods” section.

We examined association with schizophrenia and markers in NRG1 and related genes in an ErbB4 signaling pathway and validated functional effects of associated SNP interactions by testing these interactions on brain function in normal subjects carrying risk-associated alleles using functional neuroimaging. We previously demonstrated that putative intermediate phenotypes at the level of brain information processing are robust reflections of biological mechanisms related to potential schizophrenia susceptibility genes.33 Our primary emphasis was to examine epistasis, focusing on genes with prior evidence for association and conditional on established biological interaction between protein products encoded by each gene. Because power to show higher-order interactions is dependent on sample size and effect size, we restricted this study to a few genes in a specific NRG1-related biological network that has been implicated in schizophrenia.34

NRF1 is a complex gene with many biological effects. Its role as a potential susceptibility factor in schizophrenia may be related to a number of these functions, but its role in NMDA signaling has been of particular interest.35,36 NRG1 pathway partners implicated in NMDA signaling were selected because they either directly interacted with neuregulin 1 (erbB437 [gene symbol, ERBB4]) and PSD-9538 [DLG4]) or they interacted directly with genes in the extended ERBB4-NMDA signaling pathway (PSD-95–nNOS139 [NOS1], nNOS1–neuronal nitric oxide synthase40 [NOS1AP], and AKT34 [AKT1]) (Figure 2). Only 1 of the partner genes selected has not shown significant independent evidence for association with schizophrenia in prior studies (DLG431,41-43) but was included because of the central role of PSD-95 in NMDA function and also because of evidence of altered coupling of ErbB4 and DLG4 in the schizophrenic brain.34

**Figure 1.** Diagram showing random forest algorithm. Before growing a classification tree, steps 1 and 2 are performed, selecting a subset of cases and controls and a subset of single-nucleotide polymorphisms (SNPs) from the full set of observations and predictors. In step 3, while the tree is being grown on the same subset of observations, different subsets of SNPs are selected at each node, allowing SNPs without strong main effects to participate in interactions. The goal of the recursive partitioning is to provide purer (ie, only cases or only controls) daughter nodes. Because this can lead to overfitting of chance fluctuations in the data, the out-of-bag sample, which is independent of the observations used to grow each tree, is used to calculate the measures of variable importance, as shown in step 4, and the decrease in misclassification of cases and controls is calculated. After permuting predictor status in this same out-of-bag set (eg, under H0), these observations are again run through the tree, and the difference in the misclassification rate for the observed vs permuted data is calculated. In step 5, we create a forest of these classification trees by repeating steps 1 to 4 5000 times. Finally, the measures of variable importance are calculated for each SNP by taking the average of the decrease in classification performance across all the trees in the forest. Figure based on Nicodemus et al.32

**Figure 2.** NRG1 interaction partners. Line color denotes method of interaction detection: red=in vivo, in vitro, and yeast 2 hybrid; blue=in vivo and in vitro; violet=in vitro; and black=competition for mutual protein binding partner.

THE CLINICAL BRAIN DISORDERS BRANCH/NATIONAL INSTITUTE OF MENTAL HEALTH SIBLING STUDY

This study was conducted according to the principles expressed in the Declaration of Helsinki and approved by the institutional review board of the National Institute of Mental Health. All patients provided written informed consent for the collection of samples and for genotyping and phenotyping in...
obtained using the TaqMan 5′ exonuclease allelic discrimination assay (Applied Biosystems, Foster City, California) (eAppendix). Supplementary Table 1 lists all markers tested.

### THE GERMAN SAMPLE

Individuals with schizophrenia (n=905) and unrelated controls (n=1323) were ascertained from Munich, Germany, and self-identified as white. The details of ascertainment, screening, and diagnosis are described elsewhere.45

### GENETIC MARKER SELECTION AND GENOTYPING

We selected 49 markers in NRG1 based on reports of association with schizophrenia in previously published studies1-14 or using tagSNPs from phase 1 of the International HapMap Project (http://www.hapmap.org). The SNPs in NOS1AP (n=31), DLG4 (n=14), and NOS1 (n=4) were based on HapMap tagSNPs; SNPs in ERBB4 (n=11) were selected on previously published associated SNPs,21 likewise for AKT1 SNPs (n=12) and HapMap tagSNPs within the associated region. TagSNPs were selected using 2-marker tagging from Haploview, based on the minimal number of possible SNPs with an r² cutoff value of 0.8 and minor allele frequency more than 5%, and extended 10 kilobases 5’ and 5 kilobases 3’ of the gene, using the May 2004 build. In both data sets, blood was collected and DNA was extracted using standard methods. Genotypes for all SNPs were determined using the TaqMan 5′ exonuclease allelic discrimination assay (Applied Biosystems, Foster City, California) (eAppendix, eTable, and Supplementary Table 1). Because permutation-based VIMs may be slightly unstable,36,67 we repeated each analysis 500 times to obtain medians. Further, analyses were performed on 500 null replicates where case status had been randomly permuted to obtain empirical P values. Following MLA analysis, we performed 2-way interaction modeling via logistic regression among the subset of SNPs considered influential by MLAs. Significance of the logistic regression models was obtained using a likelihood ratio test (LRT) comparing nested models, a reduced model containing main effect terms vs a full model containing main effect and interaction terms.

### Forest-Based Algorithms

Both RF27 and CIF29,30 rely on multiple classification trees. To begin, a subset of predictors (n=10) is randomly sampled. In addition, a subsample (63.2%) of the observations is selected for tree building.29 A single tree is created using recursive partitioning of the subtree of predictors on the subsampled observations. Random forest terminates when no additional variables produce significant decreases in impurity or when the terminal node size is less than 10 observations; the stopping rule in CIF uses the multiplicity-corrected P value of the χ² test. This process is repeated to create a forest of classification trees (N=5000). The VIMs are calculated using the 36.8% left-aside independent observations; the observed mean decrease in accuracy of prediction of these independent observations per predictor vs that observed when predictor status has been permuted is averaged across all the trees in the forest.

### Monte Carlo Logic Regression

Monte Carlo logic regression31 starts with a single tree, then allows several moves during modeling, which is conducted via an rjMCMC (reversible jump Markov chain Monte Carlo) algorithm to find the best-fitting tree. Because MCLR accepts binary predictors, we recoded each SNP into 2 variables: 1 with the minor allele coded as dominant and 1 as recessive. Monte Carlo logic regression used a burn-in interval of 10 000 and a Markov chain length of 1 million. For consistency with RF/CIF, the maximum model size was set to 2 trees/4 leaves (8 predictors). The VIM is a count of the number of times a predictor is selected into a tree.

### FUNCTIONAL MAGNETIC RESONANCE IMAGING

#### Subjects

Two samples of controls participated in 2 functional magnetic resonance imaging (fMRI) studies designed to explicitly test the effect of NRG1 intragenic epistasis and NRG1/ERBB4/AKT1 epistasis on the physiological blood oxygen level–dependent response in the brain during an N-back working memory task. The 2 imaging samples were largely independent, overlapping to a small extent (28 of 172). Within imaging samples, all SNPs were in HWE. For the NRG1 intragenic epistasis study (rs4560751 and rs3802160), 87 subjects were included (eTable). The subgroups by rs4560751 and rs3802160 (6 cells) did not differ by age, IQ, or 2-back accuracy or reaction time. This allowed for a comparison of the genetic effect on brain physiology per se, ie, how the brain handles the information without confounding by task performance. The genotype groups differed by sex: the subgroup rs4560751 heterozygous (1/2) and rs3802160 homozygous for the minor allele (2/2) (n=4) consisted of males only. Consequently, we added

©2010 American Medical Association. All rights reserved.
sex as a covariate in the fMRI design to adjust for any potential sex effect. For the NRG1/ERBB4/AKT1 epistasis study (rs10503929, rs1026882, and rs2494734), 114 subjects were included. In this sample, there were no main effects or interactions of these 3 SNPs on age, sex, or IQ during behavioral performance on task.

Task and Functional Image Processing

During the fMRI acquisition, the participants performed the N-back task as previously described. Whole-brain blood oxygen level-dependent fMRI data were acquired on a GE Signa 3-T scanner (GE Systems, Milwaukee, Wisconsin) with a gradient-echo echoplanar imaging pulse sequence acquisition (24 contiguous axial slices of dimensions 3.75 × 3.75 × 6 mm; flip angle, 90°; repetition time, 2000 milliseconds; echo time, 30 milliseconds; field of view, 24 cm; matrix, 64 × 64 voxels). Images were processed with SPM5 software (http://www.fil.ion.ucl.ac.uk/) with realignment and correction for movement artifacts, spatial normalization in a standard stereotactic space (Montreal Neurological Institute template), and smoothing with a 10-mm full-width half-maximum gaussian filter. First-level images for each subject were created by modeling the 2 experimental conditions (2B and 0B) as boxcarts convolved with a canonical hemodynamic response. A contrast image for the 2B greater than 0B contrast was estimated for each subject. These contrast images were used for a second-level random-effect analysis. It was not possible to partition each genotype group across these epistatic analyses based on sex; thus, we included sex as a covariate in all subsequent analyses. Additionally, we and other groups have shown that “risk” alleles generate dorsolateral prefrontal cortex (DLPFC) inefficiency; thus, we used a DLPFC region of interest and 1-sided t tests where this directionality could reasonably be predicted for the interactions, as indicated later (eAppendix).

Second-Level Image Analysis

To study the epistatic effect of rs4560751 × rs3802160, a full factorial model in SPM5 (2-way analysis of variance) with 2 factors, rs4560751 (2 levels) and rs3802160 (3 levels), was used. Correspondingly, 6 cells were created: rs4560751 (homozygous for the major allele [1/1]) and rs3802160 (1/1, 1/2, and 2/2) and rs4560751 (1/2) and rs3802160 (1/1, 1/2, and 2/2). For rs4560751, there were no minor allele homozygotes in our sample. Since we were specifically interested in the directionality of blood oxygen level-dependent response intensity corresponding to the pattern of inefficiency on the background of risk alleles in the 2 SNPs, we tested this hypothesis by 1-tailed t contrasts. To examine the effect of NRG1 (rs10503929) and ERBB4 (rs1026882) interaction, a 2-way analysis of variance was constructed based on the risk architecture suggested by the clinical association, with individuals carrying homozygous major alleles considered at risk, in contrast to the minor allele carriers of each SNP. Significant brain activation was defined surviving P < .05 corrected for number of voxels in this independently selected region of interest. Finally, to examine the 3-SNP interaction of NRG1 and ERBB4 SNPs with AKT1 rs2494734, parameter estimates were extracted from the region of interest peak showing the 2-way NRG1 × ERBB4 effect and conditioned on the AKT1 effect predicted by the clinical association in a 3-way analysis of variance model, with significance set at P < .05 in this higher-order analysis. We used a nested analysis of variance design with 3 factors: gene (NRG1, ERBB4, and AKT1), genotype (1/1, 1/2, and 2/2), and risk nested within genotype and gene (1/1 = risk genotype for NRG1 and ERBB4 and 1/2/2/2 = risk genotype for AKT1; 1/2/2/2 = nonrisk genotype for NRG1 and ERBB4 and 1/1 = nonrisk genotype for AKT1). This secondary analysis was conducted in 3 steps: We compared the nonrisk group (no risk genotypes) with each of the groups carrying the risk genotype in only 1 NRG1, ERBB4, or AKT1 gene; the nonrisk group vs groups carrying the risk genotypes in 2 and 3 risk genotypes; and the group with 3 risk genotypes vs all the other subjects collapsed in 1 group. We used 1-tailed t contrasts for these comparisons and we report results surviving a statistical threshold of P < .05 corrected within the predefined DLPFC regions of interest according to the principles of gaussian random fields.

RESULTS

SINGLE-GENE ANALYSES

In NRG1, SNPs NRG221132, rs4560751, and rs327417 showed a trend toward deviation from HWE in controls (P values = .043, .036, and .0085, respectively), and in cases, rs3802158, rs3802160, and rs1023911 were out of HWE (P values = .027, .0075, and .046). Although both cases (P value = .018) and controls (P value = .035) showed deviation from HWE for rs1948098, this is likely due to strong association between this SNP and case status, so it was retained. Further, recent studies have shown that although SNPs out of HWE may be less powerful, they do not appear to increase false-positive findings. Neither HapICE nor any single SNP within HapICE was significantly associated with schizophrenia (Supplementary Table 1). However, positive association was observed between the +2.0 allele at the 478B14-848 NRGl microsatellite and schizophrenia (OR, 1.72; 95% confidence interval [CI], 1.16-2.56; P value = .007) and for several SNPs’ 3’ to HapICE (Supplementary Table 1 and Figure 3). Four SNPs in NOS1AP showed evidence for association with schizophrenia (Supplementary Table 1) as did AKT1 SNP rs1130233. No significant association was observed in ERBB4, NOS1, or DLG4. None of the markers showed significant association with schizophrenia after correction for multiple testing.

MLA RESULTS

In the first MLA analyses, 7 SNPs had median VIM rankings in the top 10% by all 3 MLAs, including 1 SNP in the HAPICE region of NRG1 (rs4560751) and 4 SNPs located 3’ to the HAPICE region (rs1948098, rs3802160, rs3802158, and rs10503929); 1 SNP in ERBB4 (rs1026882); and 1 SNP in NOS1AP (rs7538490). The empirical P values based on 500 null reiterations revealed that only rs1026882 in ERBB4 was not significantly associated at the .05 level, although P values estimated using MCLR and RF were suggestive (P values = .094 and .052). Because AKT1 is regulated by the upstream effects of the genes in the first analysis, a secondary analysis was performed adding AKT1, which produced virtually identical results as in Table 1 and also highlighted a SNP in AKT1, rs2494734 (P values = .018 and <.002 for RF and CIF).

We confirmed epistasis and estimated the effect size of 2-SNP interactions using logistic regression. Of the total 20 interactions performed (C2 = 21; 1 model could not be tested because of sparse cells), 4 were significant at the .05 level (expected = 1), and 2 of the LRT P values testing intragenic interaction in NRG1 (rs4560751/rs3802160 and
rs4560751/rs3802158) passed Bonferroni correction for 20 tests. In all, interactions were observed between rs4560751 and three 3′SNPs in NRG1 and between the NRG1 3′ non-synonymous SNP rs10503929 and the ERBB4 SNP rs1026882 (Table 2). Because AKT1 is downstream of and regulated by NRG1/ERBB4 interactions, we specifically tested whether a 3-way interaction between NRG1 SNP rs10503929, ERBB4 rs1026882, and AKT1 rs2494734 was present. In comparison with the reference group (25 controls; 14 cases), who carried no risk genotypes at any of the 3 loci, we found 3.87-fold (95% CI, 1.20-12.44; P value=.023; 6 controls; 13 cases) increased risk for individuals carrying only the risk genotype at AKT1 rs2494734; a 2.38-fold (95% CI, 1.11-5.09; P value=.025; 45 controls; 60 cases) increased risk for those carrying only the risk genotype at NRG1 rs10503929, the same NRG1/ERBB4 interaction as mentioned earlier (OR, 3.27; 95% CI, 1.37-7.81; P value=.0080; 18 controls; 33 cases); and a 27.13-fold (95% CI, 3.30-223.03; P value=.0020; 5 controls; 17 cases) increased risk for schizophrenia among individuals carrying all 3 risk genotypes vs those carrying no risk genotypes (LRT P value=.042; P value for model=.0082), suggesting the AKT1 interaction may be dependent on the NRG1/ERBB4 interaction. AKT1 did not statistically interact directly with ERBB4 or NRG1, which is interesting in relation to what is known about its biology.36 No other genotype combination was statistically different from the reference group. Although cell sizes were all at least 5, the wide CIs indicate imprecision in estimation and warrant cautious interpretation. In the German case-control sample, we observed a marginally significant interaction (LRT P value=.054) between NRG1 SNPs rs4560751 and rs3802158; however, the risk allele at rs3802158 was the opposite of the original sample.

NEUROIMAGING RESULTS

To extend our genetic findings and test their validity at the level of human brain function, we tested whether epistatic effects on risk would show similar effects on aspects of brain function associated with increased risk for illness. In other words, if interactions between specific SNPs in these genes increase risk for schizophrenia, this happens presumably because the interaction increases the expression of biological abnormalities in the brain that also are associated with increased risk of manifest illness. We focused this level of analysis on a brain-based intermediate phenotype relevant for schizophrenia and human cognitive function,48 a pattern of inefficient activation in DLPFC revealed with an fMRI working memory paradigm (the N-back task) in normal subjects carrying risk-associated alleles. The fMRI efficiency response during this task has been shown to be heritable,51 and increased activation for a fixed level of performance has been associated with increased genetic risk for schizophrenia48.
and with genes implicated as schizophrenia susceptibility genes.33 We tested 2 specific biological hypotheses from the clinical data on the blood oxygen level–dependent physiological response in DLPFC during the N-back task48: epis-

tasis between rs4560751 and rs3802160 within NRG1 (we tested only 1 of the 3 NRG1 intragenic models because 3 SNPs in the 3’ end of NRG1 were in linkage disequilibrium) and 2- and 3-SNP epistasis between rs10503929, rs1026882, and rs2494734 (NRG1, ERBB4, and AKT1). For image presentation, a statistical parametric map was overlaid onto the “single-subject T1” canonical image within SPM5 and the image was thresholded at P < .05 uncorrected for display purposes only.

**EPISTATIC EFFECT OF NRG1**

**RS4560751/RS3802160 ON DLPFC FUNCTION**

Consistent with the epistasis observed in the case-control sample, an interaction between rs4560751 and rs3802160 was observed in DLPFC (Brodman area [BA] 10) (Talairach coordinates x, y, and z: -27, 48, and 20; z score, 3.07; P = .025 familywise error [FWE] corrected) (Figure 4). Carriers of risk alleles at both genotypes revealed the most inefficient activation. On the background of rs4560751 major allele homozygosity, the mean blood oxygen level–dependent response increased in a

**lele dose-dependent fashion for rs38021160 (1/1<1/2<2/2)** (Figure 4), while on the nonrisk rs4560751 background, the opposite effect was observed.

**EPISTATIC EFFECTS OF NRG1/ERBB4/AKT1 SNPS ON DLPFC FUNCTION**

We investigated the combined physiological effects of NRG1, ERBB4, and AKT1. A significant interaction of NRG1 rs10503929 and ERBB4 rs1026882 was observed at the DLPFC, where individuals who were homozygous for risk SNPs in both NRG1 and ERBB4 were disproportionately inefficient in cortical processing (peak Talairach coordinates x, y, and z: 42, 39, and 33; F1,110=11.308; z=3.20; P < .001 uncorrected; P < .05 FWE corrected) (Figure 5). To explicitly test the prediction that conditioning on the AKT1 SNP would further impact the risk architecture of the NRG1 and ERBB4 SNPs on DLPFC function, as suggested by the 3-way clinical interaction, we performed a 3-way analysis of variance on extracted parameter estimates from this same peak, which showed that the NRG1/ERBB4 epistasis was accentuated in the background of homoyzgous AKT1 risk alleles. Individuals who carried all 3 NRG1/ERBB4/AKT1 risk genotypes were disproportionately inefficient in DLPFC function at this location (F1,199=3.663; P = .35) (Figure 6). A subsequent analysis first compared DLPFC function in

---

**Table 1. Median Variable Importance Values and Associated Empirical P Values From 3 Machine Learning Algorithms**

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>SNP</th>
<th>Observed Range</th>
<th>Median</th>
<th>Permute Range a</th>
<th>Median</th>
<th>Permute Range a</th>
<th>Median</th>
<th>Permute Range a</th>
<th>Median</th>
<th>Permute Range a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>rs4560751</td>
<td>0.59 to 1.07</td>
<td>0.835</td>
<td>-0.45 to 1.17</td>
<td>.01</td>
<td>-2.4 x 10^-4</td>
<td>0.0012</td>
<td>-1.2 x 10^-3</td>
<td>0.024</td>
<td>4.0 to 13.2</td>
</tr>
<tr>
<td></td>
<td>rs1948098</td>
<td>0.98 to 1.45</td>
<td>1.23</td>
<td>-0.49 to 0.95</td>
<td>.002</td>
<td>4.8 x 10^-4</td>
<td>0.0021</td>
<td>1.2 x 10^-3</td>
<td>0.004</td>
<td>54.2 to 65.1</td>
</tr>
<tr>
<td></td>
<td>rs3802160</td>
<td>0.34 to 0.83</td>
<td>0.56</td>
<td>-0.49 to 0.81</td>
<td>.012</td>
<td>-2.3 x 10^-4</td>
<td>0.0012</td>
<td>2.8 x 10^-3</td>
<td>0.028</td>
<td>2.3 to 11.9</td>
</tr>
<tr>
<td></td>
<td>rs3802158</td>
<td>0.19 to 0.71</td>
<td>0.49</td>
<td>-0.48 to 0.76</td>
<td>.024</td>
<td>-2.8 x 10^-3</td>
<td>0.001</td>
<td>2.9 x 10^-3</td>
<td>0.028</td>
<td>1.0 to 7.1</td>
</tr>
<tr>
<td></td>
<td>rs10503929</td>
<td>0.22 to 0.69</td>
<td>0.5</td>
<td>-0.48 to 1.28</td>
<td>.018</td>
<td>-1.6 x 10^-4</td>
<td>0.0111</td>
<td>9.2 x 10^-3</td>
<td>0.03</td>
<td>4.0 to 7.1</td>
</tr>
<tr>
<td></td>
<td>ERBB4 rs1026882</td>
<td>0.01 to 0.55</td>
<td>0.28</td>
<td>-0.49 to 0.83</td>
<td>.004</td>
<td>-1.1 x 10^-3</td>
<td>0.00054</td>
<td>-1.3 x 10^-3</td>
<td>0.13</td>
<td>2.6 to 4.2</td>
</tr>
<tr>
<td></td>
<td>NOS1AP rs7538490</td>
<td>0.20 to 0.73</td>
<td>0.45</td>
<td>-0.43 to 1.15</td>
<td>.018</td>
<td>-4.4 x 10^-4</td>
<td>0.00009</td>
<td>1.2 x 10^-2</td>
<td>0.036</td>
<td>5.8 to 8.8</td>
</tr>
<tr>
<td></td>
<td>AKT1 rs2494734</td>
<td>0.27 to 0.77</td>
<td>0.52</td>
<td>-0.49 to 2.04</td>
<td>.018</td>
<td>-4.8 x 10^-3</td>
<td>0.0081</td>
<td>9.9 x 10^-4</td>
<td>&lt;.002</td>
<td>0.6 to 1.9</td>
</tr>
</tbody>
</table>

**Abbreviation: SNP, single-nucleotide polymorphism.**

**Table 2. Significant 2-SNP Interactions Between SNPs Selected via Machine Learning Algorithms**

<table>
<thead>
<tr>
<th>Interaction</th>
<th>SNP 1</th>
<th>SNP 2</th>
<th>SNP 1 Risk Genotype</th>
<th>SNP 2 Risk Genotype</th>
<th>Interaction OR (95% CI)</th>
<th>LRT P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRG1 x NRG1</td>
<td>rs4560751</td>
<td>rs1948098</td>
<td>G/G</td>
<td>A/A</td>
<td>1.25 (0.66-2.39)</td>
<td>.016</td>
</tr>
<tr>
<td>NRG1 x NRG1</td>
<td>rs4560751</td>
<td>rs3802160</td>
<td>G/G</td>
<td>T carrier</td>
<td>4.56 (1.97-10.58)</td>
<td>.00020</td>
</tr>
<tr>
<td>NRG1 x NRG1</td>
<td>rs4560751</td>
<td>rs3802158</td>
<td>T carrier</td>
<td>G carrier</td>
<td>0.29 (0.13-0.61)</td>
<td>.00010</td>
</tr>
<tr>
<td>NRG1 x AKT1</td>
<td>rs10503929</td>
<td>rs1026882</td>
<td>C/C</td>
<td>A/A</td>
<td>2.25 (1.11-4.54)</td>
<td>.035</td>
</tr>
</tbody>
</table>

**Abbreviations: CI, confidence interval; LRT, likelihood ratio test; OR, odds ratio; SNP, single-nucleotide polymorphism.**
groups carrying 1 risk genotype at any of the 3 genes with the nonrisk genotype carriers in all genes; none showed significant DLPFC inefficiency. We then compared groups carrying risk genotypes at NRG1/AKT1 and ERBB4/AKT1 and all 3 risk genotypes with the nonrisk genotype carriers. While none of the groups with 2 risk genotypes significantly differed from the nonrisk group, the carriers of risk genotypes in all 3 genes (n=4) combined demonstrated significantly greater activation compared with nonrisk carriers (n=17) in the right middle frontal gyrus (BA 46; Talairach coordinates x, y, and z: 45, 39, and 26; z score = 3.19; P = 0.019 FWE corrected) (Figure 6A). The group with all 3 risk genotypes was also more inefficient compared with the group with 2 risk genotypes in NRG1 and AKT1 (BA 9; Talairach coordinates x, y, and z: 45, 39, and 28; z score = 4.53; P = 0.0002 FWE corrected; n=12), the group with 2 risk genotypes in ERBB4 and AKT1 (BA 46, Talairach coordinates x, y, and z: 45, 39, and 26; z score = 3.01; P = 0.03 FWE corrected; n=3), and all other subjects at the right DLPFC (BA 9; Talairach coordinates x, y, and z: 42, 39, and 31; z score = 3.23; P = 0.015 FWE corrected; n=95) (Figure 6B). These findings support the clinical results that showed association with case status for the risk carriers in all the 3 genes but not for the risk carriers in NRG1/AKT1 or ERBB4/AKT1. A post hoc statistical analysis of this contrast (carriers of all 3 risk genotypes vs those carrying any other combination) in the schizophrenia case-control sample revealed a 3.04-fold increase in risk (95% CI: 1.09-8.45; P value = 0.033) for those carrying all 3 risk-associated genotypes. Thus, each of the SNP interactions observed in the clinical data sets impacting risk for schizophrenia showed consistent effects on cortical function in an independent sample of normal individuals carrying combinations of the same SNP genotypes.

CONFIRMATORY ANALYSES IN PUBLIC GENOME-WIDE ASSOCIATION DATASETS

We searched a schizophrenia genome-wide association study data set (GAIN) (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000021.v2.p1)23 and the data set described by Need et al13 for replication of epistasis. The GAIN data set consists of 1172 cases with schizophrenia or schizoaffective disorder and...
1378 controls and was genotyped on the Affymetrix 6.0 platform (Affymetrix, Santa Clara, California). The Need et al data set comprised 2 separately collected data sets: 1 using the Illumina HumanHap300 chip (Illumina, San Diego, California) from Germany (630 cases, 547 controls) and 1 using the Illumina HumanHap550 chip from Aberdeen, Scotland (669 cases, 597 controls); cases in both studies had to qualify for diagnosis of schizophrenia under both DSM-IV and International Statistical Classification of Diseases, 10th Revision. None of our VIM SNPs from MLA analyses were genotyped in the GAIN sample. Imputation of these SNPs in GAIN could not be performed because only 1 SNP participating in epistasis was found in HapMap. Therefore, we tested the closest flanking SNPs in the GAIN sample, assuming these would be among the most likely to show linkage disequilibrium with our SNPs or with an unknown causal variant. Using the GAIN data and flanking SNPs for our 5’ and 3’ epistasis SNPs within NRG1 (a total of 6 SNPs, after performing 8 possible 2-SNP interaction tests), we observed 2 interactions with LRT $P$ values less than .05 (rs7812662 × rs7830691 [.024] and rs10098401 × rs17631978 [.042]), uncorrected, and 2 interaction models with suggestive evidence for interaction ($P$ values < .10: rs7812662 × rs6468119 [.079] and rs7812662 × rs17631978 [.094]); none of the 6 SNPs were individually associated ($P$ values > .10). The expected number of tests showing $P$ values less than .05 for 8 independent tests is 0.4 and the number expected to show $P$ values less than .10 is 0.8; we observed 2 and 4, respectively. Thus, in 4 independent samples reported (the present study, German, and GAIN samples and the sample reported in Benzel et al), modest evidence for epistasis between SNPs in the 5’ and 3’ regions of NRG1 has been observed (eAppendix, eTable, and Supplementary Table 1). Because of low minor allele frequency at the closest flanking ERBB4 SNPs in this data set, we were unable to test the 2- and 3-way interaction between NRG1, ERBB4, and AKT1 (eAppendix, eTable, and Supplementary Table 1).

In the data described by Need et al., consisting of an independent sample from Aberdeen and an overlapping sample from Germany to our German samples, the NRG1 and ERBB4 SNPs participating in the 3-way interaction were genotyped; however, additional SNPs were not and, as earlier, were not genotyped in HapMap. The SNPs genotyped in the Need et al. data overlapped with SNPs genotyped in our data; thus, proxies were selected using link-
age disequilibrium from our control sample for the intragenic NRG1-interacting SNPs and the AKT1 SNP. While there were trends for association with schizophrenia at 4 of the 5 single SNPs tested, we did not replicate our interactions in this pooled data set because of allelic heterogeneity of the SNPs selected and very modest linkage disequilibrium between the SNPs selected as proxies and our SNPs (eAppendix, eTable, and Supplementary Table 1). However, in the separate data sets, individuals carrying all 3 risk genotypes (using the Aberdeen coding of risk genotypes at ERBB4) did show marginally increased risk for schizophrenia (OR, 1.60; 95% CI, 0.98-2.61; P value = .062) vs those carrying no risk genotypes in the German sample; the same pattern of results was found in the Aberdeen sample (LRT P value = .076), with individuals carrying all 3 risk genotypes showing increased risk (OR, 1.91; 95% CI, 0.98-3.69; P value = .056). The NRG1 5′ and 3′ interaction was not replicated in the Aberdeen sample, possibly because of the low linkage disequilibrium between the proxy SNPs and our epistasis SNPs (r² = 0.13 and 0.30).

**COMMENT**

Epistasis between SNPs within NRG1 and between NRG1 and selected protein interaction partners influenced risk for schizophrenia, which we biologically validated using fMRI in healthy controls (Figure 7). Individuals carrying the same combinations of schizophrenia risk genotypes showed significantly less efficient cognitive processing, similar to findings in patients with schizophrenia and in their healthy siblings. A 3-way interaction between a nonsynonymous SNP in NRG1 and SNPs in ERBB4 and AKT1 was associated with substantial increased risk for schizophrenia and inefficient physiological processing of working memory in normal subjects. Markers in NRG1 showing epistasis were found at the 5′ and 3′ ends of the gene in regions that have been individually associated in earlier studies,1-12 and we also observed interaction between 1 marker in the 5′ end and 3 markers at the 3′ end confirming previous reports.24,25 The SNPs in AKT1/ERBB4 interacted with the nonsynonymous NRG1 SNP Thr286/289/294Met (rs10503929), which has been reported to be marginally associated with schizophrenia,8,19 and a recent meta-analysis of this SNP showed statistically significant association.33 Epistatic interactions were stronger than single-SNP effects and in some cases occurred between SNPs not showing main effects, a classic characteristic of epistasis described in other genetic contexts.34

The MLA P values were generated using experiment-wise permutation testing and neuroimaging analyses used conservative FWE correction.35 However, we have performed a number of statistical tests, and we report uncorrected P values for analyses using unconditional logistic regression. We recognize that few of our P values for these regression models would survive naïve correction for all tests. Nevertheless, we have identified evidence within the GAIN data set of interactions with 5′ and 3′ SNPs in NRG1. Higher-order interactions were not testable in this data set and replication in another sample was complicated by allelic heterogeneity. Also, 4 of 5 SNPs in this latter sample showed single-SNP-level association in 1 or both of the data sets.

We have taken an alternative strategy to validate our statistical associations by hypothesis testing predetermined interactions at the level of brain function, arguing that for interactions of SNPs to exaggerate clinical risk, they should exaggerate the biological underpinnings of risk.36 Schizophrenia is associated with abnormal frontal lobe function, which has been extensively documented,37 and qualitatively similar abnormalities are found in healthy family members of patients with schizophrenia, implicating frontal cortex functional abnormalities as intermediate biological phenotypes related to genetic risk for schizophrenia.36,37 We used a measure of frontal lobe physiological function—the efficiency of prefrontal cortical engagement during working memory—as an intermediate phenotype to biologically validate statistical associations with clinical illness. Our fMRI results establish brain-functional interactions between the SNPs tested, implicating a neural systems mechanism for the clinical epistasis.

Schizophrenia is highly heritable, but thus far many genome-wide association studies have not been “successful” in the sense of finding genome-wide significant results that are replicated in independent studies. However, nearly all of these studies have focused on the analysis of single SNPs, which seem to fall short of explaining the heritability, with the exception of a recent study that proposed a polygenic model of schizophrenia.36 Genetic variation exists within a biological context and this con-
text is likely epistatic and/or polygenic. Studies that look for individual variant effects may be compromised in finding association without taking these contextual issues into account. Future studies will determine whether the epistatic and/or polygenic model is the rule, rather than the exception. We believe our approach of independent biological validation provides strong support that this is fertile territory for further research related to genetic risk for schizophrenia.

In conclusion, we report evidence for statistical epistasis between SNPs in the 3’ end of NRG1 with 5’ NRG1 SNPs, between NRG1 and ERBB4, and a 3-way interaction between a functional SNP in NRG1 with ERBB4 and AKT1. The 3’ end of NRG1 is exon rich, providing biological support for 5’ promoter × 3’ interactions observed in the present study and by Benzel et al., and may play a role in increasing risk for schizophrenia. The NRG1 × ERBB4 × AKT1 interaction is consistent with the work of Hahn et al., showing AKT1 function is modulated by NRG1 × ERBB4 interactions. We have validated each of these genetic interactions in terms of functional influence on cognitive processing in the prefrontal cortex of healthy controls.

Submitted for Publication: July 22, 2009; final revision received March 3, 2010; accepted April 7, 2010.

Author Affiliations: Genes, Cognition, and Psychosis Program, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland (Drs Nicodemus, Law, Radulceas, Kolachana, Vakkalanka, Straub, Callicott, Tan, and Weinberger and Mr Luna); Wellcome Trust Centre for Human Genetics (Dr Nicodemus), Department of Clinical Pharmacology (Dr Nicodemus), and Department of Psychiatry, Warneford Hospital (Dr Law), University of Oxford, Oxford, England; Section of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig Maximilians University, Munich, Germany (Drs Rujescu and Giegling); Laboratory of Experimental Immunology, National Cancer Institute, Frederick, Maryland (Drs McGee, Gold, and Dean); and GlaxoSmithKline, Verona, Italy (Dr Muglia).

Correspondence: Daniel R. Weinberger, MD, Genes, Cognition, and Psychosis Program, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, Room 4S-235, 10 Center Dr, Bethesda, MD 20892 (weinberd@mail.nih.gov).

Author Contributions: Dr Weinberger had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Financial Disclosure: Dr Muglia is a full-time employee of GlaxoSmithKline, who have filed patent applications for SNPs associated with schizophrenia (US patent applications 20080176239 and 20080176240 and international application number PCT/EP2008/050477). The recruitment of the German patients was partially supported by GlaxoSmithKline.

Funding/Support: This work was supported by the National Institute of Mental Health Intramural Research Program. Dr Law is a Medical Research Council UK Career Development Fellow.

Role of the Sponsor: The funders (National Institute of Mental Health Intramural Research Program) had no role in the design or conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript.

Online-Only Material: The eAppendix and eTable are available at http://www.archgenpsychiatry.com. Supplementary Table 1 is available at http://intramural.nimh.nih.gov/gcap/.

Additional Information: This study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health (http://biowulf.nih.gov).

REFERENCES


