

# 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 rs4560751 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.

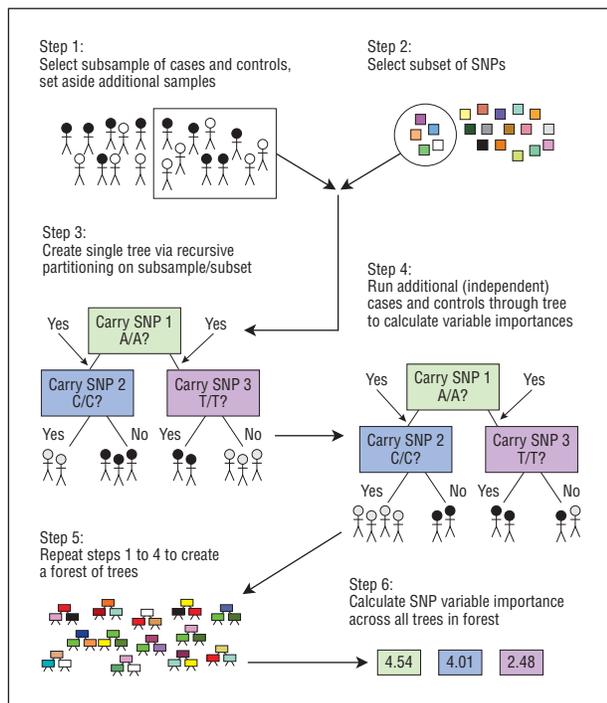
**Conclusion:** Our data suggest complex epistatic effects implicating an *NRG1* molecular pathway in cognitive brain function and the pathogenesis of schizophrenia.

*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,<sup>1–12</sup> 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,<sup>1–12</sup> although null studies have appeared.<sup>13–20</sup> One meta-analysis<sup>21</sup> of the original risk-associated haplotype first identified in an Icelandic sample (HAP<sub>ICE</sub>) revealed a small effect size (odds ratio [OR], 1.22) but strong statistical associa-

tion ( $P$  value =  $8.0 \times 10^{-10}$ ), and another showed a significant, but weaker, association ( $P$  value = .036).<sup>22</sup> 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 HAP<sub>ICE</sub> and its receptor *ERBB4*.<sup>23–25</sup> 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.<sup>26</sup>

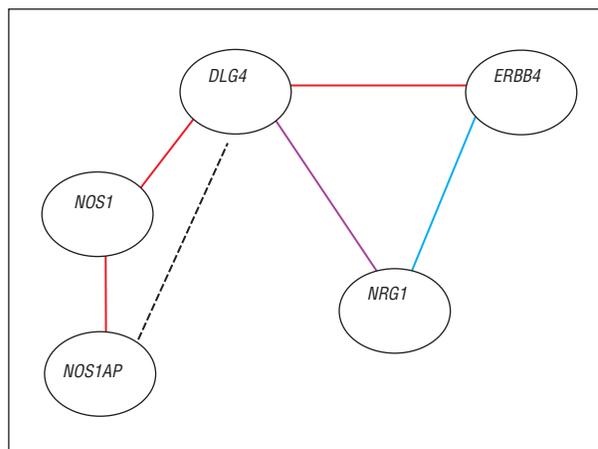
Author Affiliations are listed at the end of this article.



**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  $H_0$ ), 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 calculated in step 4 across all the trees in the forest. Figure based on Nicodemus et al.<sup>32</sup>

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 algorithms<sup>27-31</sup> 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



**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.

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.<sup>33</sup> 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.<sup>34</sup>

*NRG1* 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.<sup>35,36</sup> *NRG1* pathway partners implicated in NMDA signaling were selected because they either directly interacted with neuregulin 1 (erbb4<sup>37</sup> [gene symbol, *ERBB4*] and PSD-95<sup>38</sup> [*DLG4*]) or they interacted directly with genes in the extended *ERBB4*-NMDA signaling pathway (PSD-95–nNOS1<sup>39</sup> [*NOS1*], nNOS1–neuronal nitric oxide synthase<sup>40</sup> [*NOS1AP*], and AKT<sup>34</sup> [*AKT1*]) (**Figure 2**). Only 1 of the partner genes selected has not shown significant independent evidence for association with schizophrenia in prior studies (*DLG4*<sup>23,34,41-43</sup>) 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.<sup>34</sup>

## METHODS

### 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-

formation. Proband met *DSM-IV* criteria for a broad diagnosis category ( $n=296$ ).<sup>44</sup> Control individuals ( $n=365$ ) were ascertained from the National Institutes of Health Normal Volunteer Office and were screened for psychiatric disorders as described elsewhere.<sup>44</sup> Controls and probands were between 18 and 60 years old, without significant medical problems or history of head trauma or recent history of drug and/or alcohol abuse, and had measured IQ (for affected individuals, premorbid IQ) higher than 70. A research psychiatrist screened and interviewed all participants using the Structured Clinical Interview for *DSM-IV*. All individuals self-identified as white.

## 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.<sup>45</sup>

## GENETIC MARKER SELECTION AND GENOTYPING

We selected 49 markers in *NRG1* based on reports of association with schizophrenia in previously published studies<sup>1-12</sup> 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,<sup>22</sup> likewise for *AKT1* SNPs<sup>43</sup> ( $n=12$ ) 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  $r^2$  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 obtained using the TaqMan 5' exonuclease allelic discrimination assay (Applied Biosystems, Foster City, California) (eAppendix and eTable [<http://www.archgenpsychiatry.com>] and supplementary Table 1 [<http://intramural.nimh.nih.gov/gcap/>]). Supplementary Table 1 lists all markers tested.

## STATISTICAL METHODS

### Single-Gene Association Analyses

Testing for Hardy-Weinberg equilibrium (HWE) was performed using the Fisher exact test. Unconditional logistic regression was used to test for genotypic association between single SNPs, the R package (<http://www.r-project.org/>) haplo.stats was used to perform HAP<sub>ICE</sub> haplotype analyses, and plotting of  $-\log_{10}(P \text{ values})$  against an  $r^2$  linkage disequilibrium heat map was performed using the R package snp.plotter and included 296 unrelated cases and 365 unrelated normal controls from the National Institute of Mental Health sample. Reported single-marker  $P$  values were not adjusted for multiple testing.

### Complex Epistasis—MLAs

We used 3 MLAs (random forest [RF], conditional inference forest [CIF], and Monte Carlo logic regression [MCLR]) to assess complex interactions to find consensus across methods. We considered a SNP for follow-up if the median variable importance measure (VIM), a measure of association, was in the top 10% of the distribution of all VIMs using more than 1 algorithm. Missing genotypes were inferred using haplo.stats for individuals who had 80%

or greater genotype data ( $n=415$ ; 220 cases, 195 controls) (eAppendix, eTable, and Supplementary Table 1). Because permutation-based VIMs may be slightly unstable,<sup>46,47</sup> 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 RF<sup>27</sup> and CIF<sup>29,30</sup> 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.<sup>30</sup> A single tree is created using recursive partitioning of the subset 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  $\chi^2$  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 regression<sup>31</sup> 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

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 or during behavioral performance on task.

### Task and Functional Image Processing

During the fMRI acquisition, the participants performed the N-back task as previously described.<sup>48</sup> 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 boxcars 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.<sup>49</sup>

## RESULTS

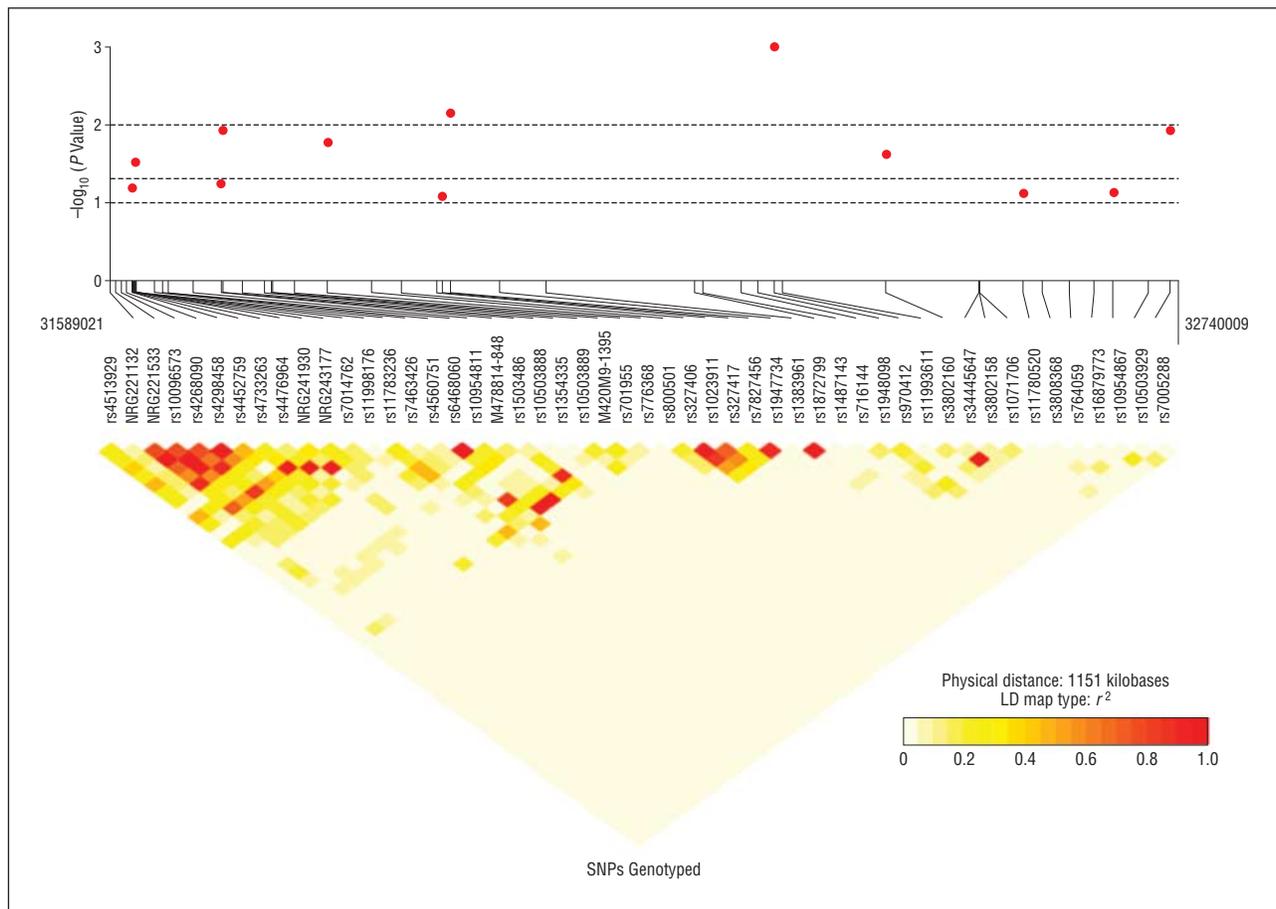
### SINGLE-GENE ANALYSES

In *NRG1*, SNP8NRG221132, rs4560751, and rs327417 showed a trend toward deviation from HWE in controls (*P* value = .043, .036, and .0085, respectively), and in cases, rs3802158, rs3802160, and rs1023911 were out of HWE (*P* value = .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.<sup>50</sup> Neither Hap<sub>ICE</sub> nor any single SNP within Hap<sub>ICE</sub> was significantly associated with schizophrenia (Supplementary Table 1). However, positive association was observed between the +2.0 allele at the 478B14-848 *NRG1* microsatellite and schizophrenia (OR, 1.72; 95% confidence interval [CI], 1.16-2.56; *P* value = .007) and for several SNPs 3' to Hap<sub>ICE</sub> (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 HAP<sub>ICE</sub> region of *NRG1* (rs4560751) and 4 SNPs located 3' to the HAP<sub>ICE</sub> region (rs1948098, rs3802160, rs3802158, and rs10503929); 1 SNP in *ERBB4* (rs1026882); and 1 SNP in *NOS1AP* (rs7538490) (**Table 1**). Empirical *P* values based on 500 null replicates 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 ( ${}^2C_2 = 21$ ; 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 interactive interaction in *NRG1* (rs4560751/rs3802160 and



**Figure 3.** Association results for *NRG1*. The  $-\log_{10}$  *P* values are plotted on the y-axis. The lines on the x-axis show the physical location of the single-nucleotide polymorphisms (SNPs) genotyped. Below the association plot is a linkage disequilibrium (LD) ( $r^2$ ) heat map showing correlation between genotyped SNPs.

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.<sup>34</sup> No other genotype combination was statistically dif-

ferent 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,<sup>48</sup> 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,<sup>51</sup> and increased activation for a fixed level of performance has been associated with increased genetic risk for schizophrenia<sup>48</sup>

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

SNP	Algorithm											
	Random Forest				Conditional Inference Forest				Monte Carlo Logic Regression			
	Observed Range	Median	Permute Range <sup>a</sup>	Median Empirical P Value	Observed Range	Median	Permute Range <sup>a</sup>	Median Empirical P Value	Range	Median	Permute Range <sup>a</sup>	Median Empirical P Value
<i>NRG1</i>												
rs4560751	0.59 to 1.07	0.835	-0.45 to 1.17	.01	-2.4 × 10 <sup>-4</sup> to 2.6 × 10 <sup>-3</sup>	0.0012	-1.2 × 10 <sup>-3</sup> to 3.0 × 10 <sup>-3</sup>	.024	4.0 to 13.2	7.4	0.34 to 12.0	.008
rs1948098	0.98 to 1.45	1.23	-0.49 to 0.95	<.002	4.8 × 10 <sup>-4</sup> to 3.4 × 10 <sup>-3</sup>	0.0021	-1.2 × 10 <sup>-3</sup> to 3.7 × 10 <sup>-3</sup>	.004	54.2 to 65.1	59.3	1.1 to 1.3	<.002
rs3802160	0.34 to 0.83	0.56	-0.49 to 0.81	.012	-2.3 × 10 <sup>-4</sup> to 3.1 × 10 <sup>-3</sup>	0.0012	-1.2 × 10 <sup>-3</sup> to 2.8 × 10 <sup>-3</sup>	.028	2.3 to 11.9	5.8	0.89 to 2.9	<.002
rs3802158	0.19 to 0.71	0.49	-0.48 to 0.76	.024	-2.8 × 10 <sup>-4</sup> to 2.5 × 10 <sup>-3</sup>	0.001	-1.3 × 10 <sup>-3</sup> to 2.9 × 10 <sup>-3</sup>	.04	2.8 to 15.4	7.8	0.38 to 46.4	.02
rs10503929	0.22 to 0.69	0.5	-0.48 to 1.26	.018	-1.6 × 10 <sup>-4</sup> to 2.6 × 10 <sup>-3</sup>	0.0011	-9.2 × 10 <sup>-4</sup> to 2.9 × 10 <sup>-3</sup>	.03	4.0 to 7.1	5	0.30 to 20.0	.026
<i>ERBB4</i> rs1026882	0.01 to 0.55	0.28	-0.49 to 0.83	.094	-1.1 × 10 <sup>-3</sup> to 2.1 × 10 <sup>-3</sup>	0.00054	-1.3 × 10 <sup>-3</sup> to 2.2 × 10 <sup>-3</sup>	.13	2.6 to 4.2	3.2	0.38 to 20.4	.052
<i>NOS1AP</i> rs7538490	0.20 to 0.73	0.45	-0.43 to 1.15	.018	-4.4 × 10 <sup>-4</sup> to 2.3 × 10 <sup>-3</sup>	0.00092	-1.2 × 10 <sup>-2</sup> to 2.3 × 10 <sup>-3</sup>	.036	5.8 to 8.8	7.3	0.36 to 14.0	.014
<i>AKT1</i> rs2494734	0.27 to 0.77	0.52	-0.49 to 2.04	.018	-4.8 × 10 <sup>-4</sup> to 2.3 × 10 <sup>-3</sup>	0.00081	-9.9 × 10 <sup>-4</sup> to 2.7 × 10 <sup>-3</sup>	<.002	0.6 to 1.9	0.95	0.46 to 90.4	.70

Abbreviation: SNP, single-nucleotide polymorphism.  
<sup>a</sup>Permute range=variable importance measure range under H<sub>0</sub>.

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

Interaction	SNP 1	SNP 2	SNP 1 Risk Genotype	SNP 2 Risk Genotype	Interaction OR (95% CI)	LRT P Value
<i>NRG1</i> × <i>NRG1</i>	rs4560751	rs1948098	G/G	A/A	1.25 (0.66-2.39)	.016
<i>NRG1</i> × <i>NRG1</i>	rs4560751	rs3802160	G/G	T carrier	4.56 (1.97-10.58)	.00020
<i>NRG1</i> × <i>NRG1</i>	rs4560751	rs3802158	T carrier	G carrier	0.29 (0.13-0.61)	.00010
<i>NRG1</i> × <i>ERBB4</i>	rs10503929	rs1026882	C/C	A/A	2.25 (1.11-4.54)	.035

Abbreviations: CI, confidence interval; LRT, likelihood ratio test; OR, odds ratio; SNP, single-nucleotide polymorphism.

and with genes implicated as schizophrenia susceptibility genes.<sup>33</sup> We tested 2 specific biological hypotheses from the clinical data on the blood oxygen level-dependent physiological response in DLPFC during the N-back task<sup>48</sup>: epistasis 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.

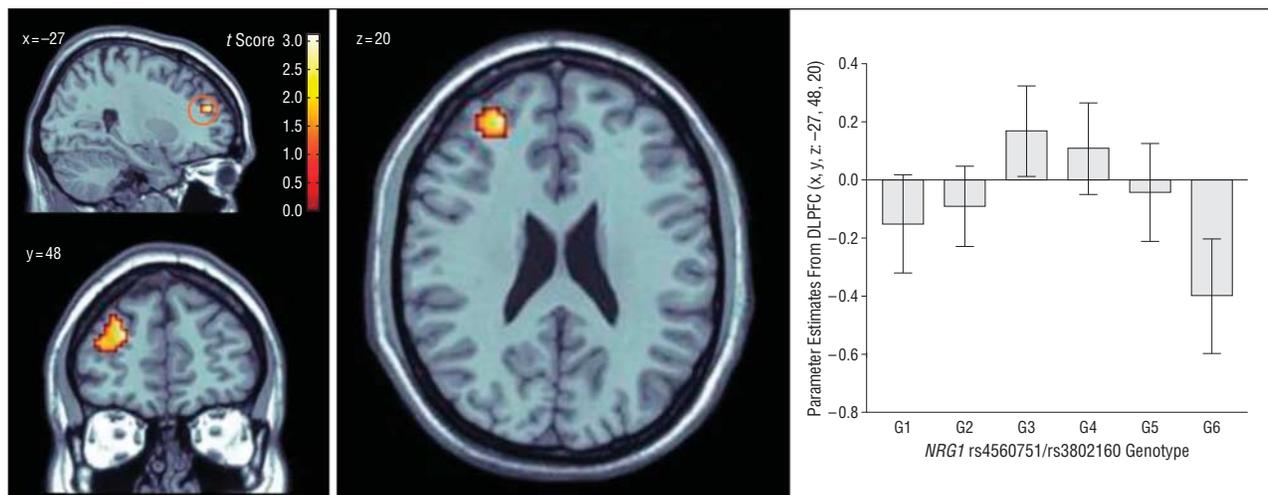
**EPISTATIC EFFECT OF *NRG1* RS4560751/RS3802160 ON DLPFC FUNCTION**

Consistent with the epistasis observed in the case-control sample, an interaction between rs4560751 and rs3802160 was observed in DLPFC (Brodmann 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 al-

lele dose-dependent fashion for rs3802160 (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;  $F_{1,110} = 11.308$ ;  $z = 3.20$ ;  $P < .001$  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 homozygous *AKT1* risk alleles. Individuals who carried all 3 *NRG1/ERBB4/AKT1* risk genotypes were disproportionately inefficient in DLPFC function at this location ( $F_{1,106} = 4.663$ ;  $P = .033$ ) (**Figure 6**). A subsequent analysis first compared DLPFC function in

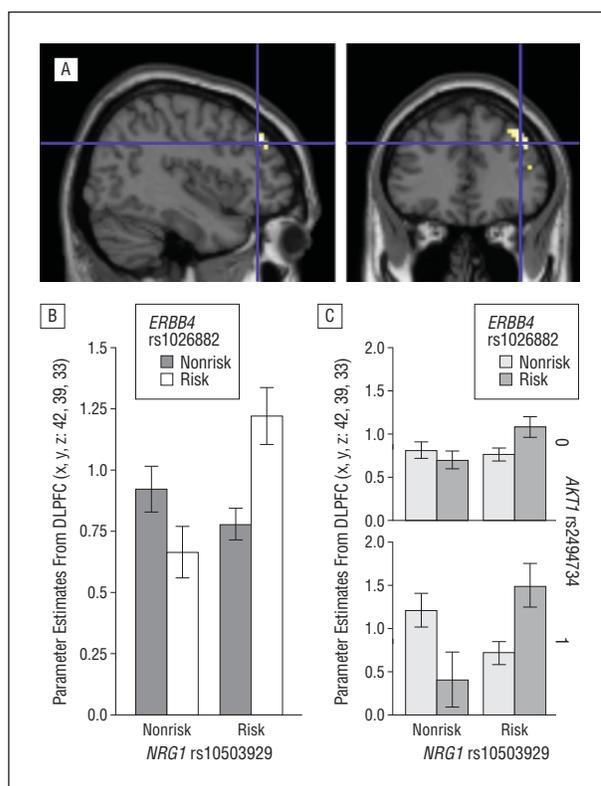


**Figure 4.** *NRG1/ERG1* neuroimaging epistasis. On the background of the risk genotype for rs4560751 (homozygous for the major allele [1/1]), the activation within left Brodmann area 10 (Talairach coordinates x, y, and z: -27, 48, and 20) increased in a dose-dependent manner for risk allele carriers with the most inefficient processing (highest activation) for the rs3802160 homozygotes. Error bars are  $\pm 1$  SE. DLPFC indicates dorsolateral prefrontal cortex; G1 group, rs4560751 1/1 and rs3802160 1/1; G2 group, rs4560751 1/1 and rs3802160 heterozygous [1/2]; G3 group, rs4560751 1/1 and rs3802160 homozygous for the minor allele [2/2]; G4 group, rs4560751 1/2 and rs3802160 1/1; G5 group, rs4560751 1/2 and rs3802160 1/2; and G6 group, rs4560751 1/2 and rs3802160 2/2.

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;  $P=.019$  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, 38, and 28;  $z$  score=4.53;  $P=.00002$  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;  $P=.03$  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;  $P=.015$  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=.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.

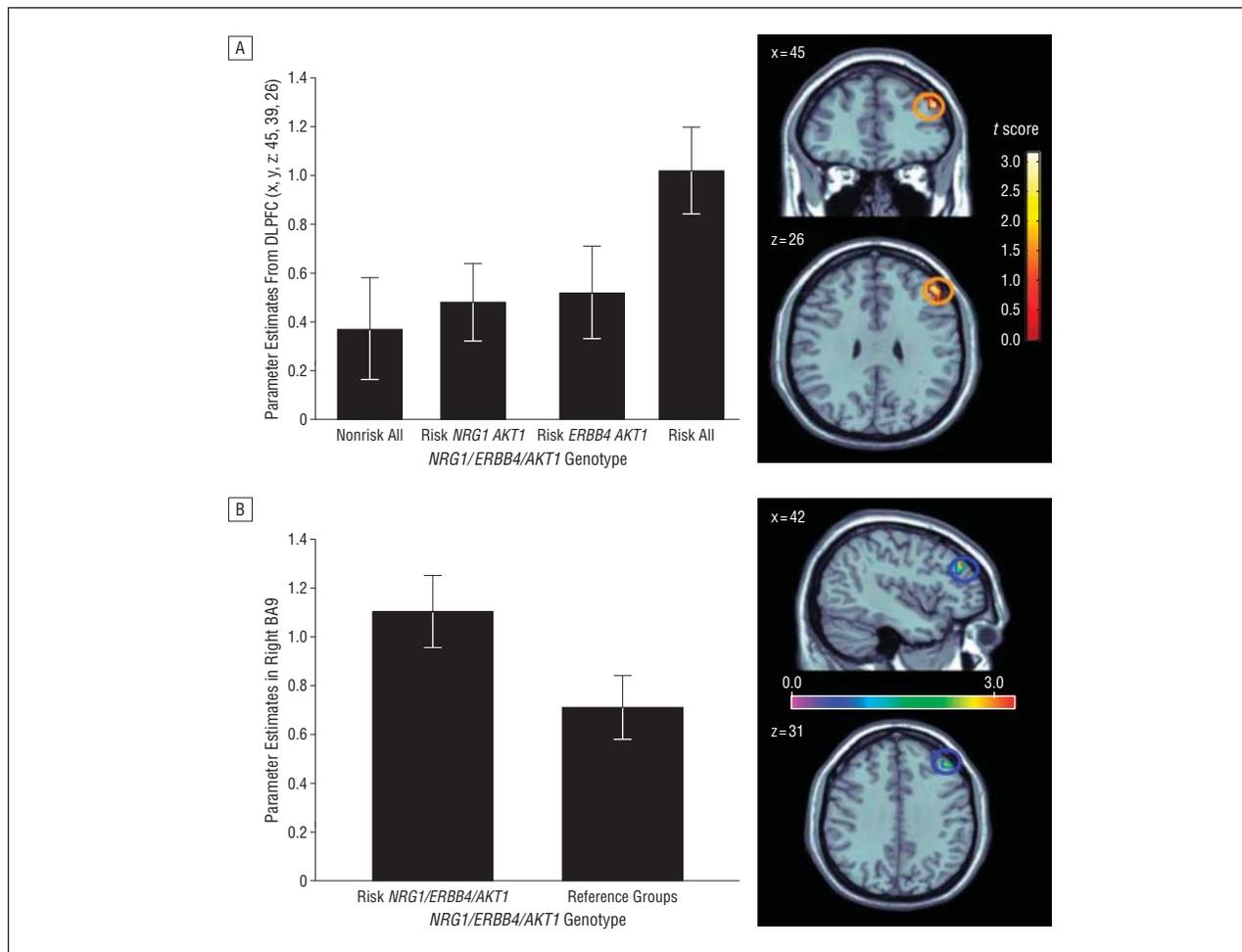
#### CONFIRMATORY ANALYSES IN PUBLIC GENOME-WIDE ASSOCIATION DATASETS

We searched a schizophrenia genome-wide association study data set (GAIN) (<http://www.ncbi.nlm.nih.gov>



**Figure 5.** *NRG1/ERBB4* neuroimaging epistasis. A, Impact of *NRG1* rs10503929  $\times$  *ERBB4* rs1026882 epistatic effect on dorsolateral prefrontal cortex (DLPFC) neural processing. B, Extracted parameter estimates from the DLPFC peak where individuals who were homozygous for the risk single-nucleotide polymorphisms in both *NRG1* and *ERBB4* were disproportionately inefficient in cortical processing. C, Individuals who carried all 3 *NRG1/ERBB4/AKT1* risk genotypes appeared disproportionately inefficient in DLPFC function at this same DLPFC location. Error bars are  $\pm 1$  SE.

/projects/gap/cgi-bin/study.cgi?study\_id=phs000021.v2.p1)<sup>52</sup> and the data set described by Need et al<sup>55</sup> for replication of epistasis. The GAIN data set consists of 1172 cases with schizophrenia or schizoaffective disorder and



**Figure 6.** *NRG1/ERBB4/AKT1* neuroimaging epistasis. A, Risk genotype carriers in *NRG1* (rs10503929; homozygous for the major allele [1/1]), *ERBB4* (rs1026882; 1/1), and *AKT1* (rs2494734; heterozygous and homozygous for the minor allele) had significantly higher activity within right Brodmann area (BA) 46 (Talairach coordinates x, y, and z: 45, 39, and 26) compared with nonrisk genotype carriers in all 3 genes. Error bars are  $\pm 1$  SE. B Subjects with risk genotypes for *NRG1*, *ERBB4*, and *AKT1* had significantly higher activity in the dorsolateral prefrontal cortex (DLPFC) (right BA 9, Talairach coordinates x, y, and z: 42, 39, and 31) during the N-back task.

1378 controls and was genotyped on the Affymetrix 6.0 platform (Affymetrix, Santa Clara, California). The Need et al<sup>45</sup> 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  $\times$  rs7830691 [.024] and rs10098401  $\times$  rs17631978 [.042]), uncorrected, and 2 interaction models with suggestive evidence for interaction

(*P* values < .10: rs7812662  $\times$  rs6468119 [.079] and rs7812662  $\times$  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<sup>25</sup>), 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).

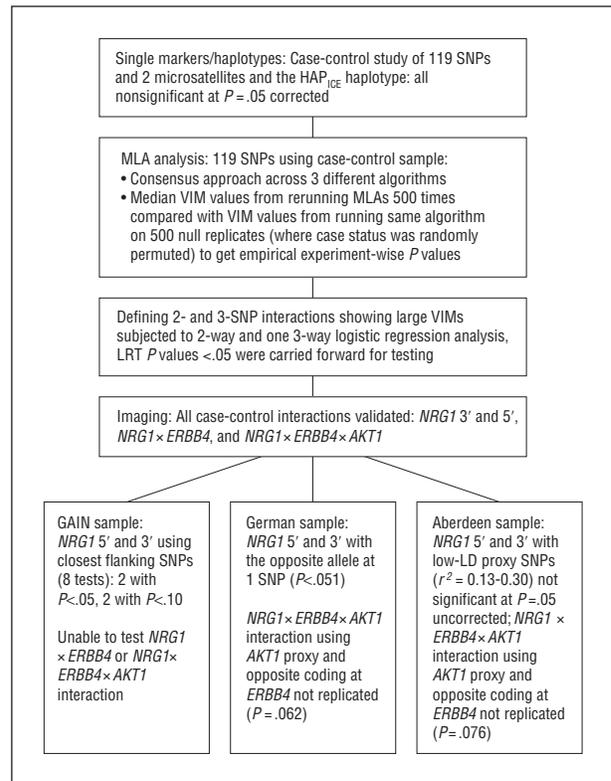
In the data described by Need et al,<sup>45</sup> consisting of an independent sample from Aberdeen and an overlapping sample from Germany to our German samples, the *NRG1* and *ERBB4* SNPs participating in the 3-way interaction were genotyped; however, additional SNPs were not and, as earlier, were not genotyped in HapMap. The SNPs genotyped in the Need et al<sup>45</sup> 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 ( $r^2 = 0.13$  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,<sup>1-12</sup> and we also observed interaction between 1 marker in the 5' end and 3 markers at the 3' end confirming previous reports.<sup>24,25</sup> 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,<sup>8,19</sup> and a recent meta-analysis of this SNP showed statistically significant association.<sup>53</sup> 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.<sup>54</sup>

The MLA *P* values were generated using experiment-wise permutation testing and neuroimaging analyses used conservative FWE correction.<sup>55</sup> 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 naive 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



**Figure 7.** Convergence of evidence for epistasis. LD indicates linkage disequilibrium; LRT, likelihood ratio test; MLA, machine learning algorithm; SNP, single-nucleotide polymorphism; VIM, variable importance measure.

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.<sup>56</sup> Schizophrenia is associated with abnormal frontal lobe function, which has been extensively documented,<sup>57</sup> 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.<sup>56,57</sup> 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.<sup>58</sup> 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  $\times$  3' interactions observed in the present study and by Benzel et al,<sup>25</sup> and may play a role in increasing risk for schizophrenia. The *NRG1*  $\times$  *ERBB4*  $\times$  *AKT1* interaction is consistent with the work of Hahn et al,<sup>34</sup> showing *AKT1* function is modulated by *NRG1*  $\times$  *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, Radulescu, 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; or 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

- Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, Brynjolfsson J, Gunnarsdottir S, Ivarsson O, Chou TT, Hjaltonson O, Birgisdottir B, Jonsson H, Gudnadottir VG, Gudmundsdottir E, Bjornsson A, Ingvarsson B, Ingason A, Sigfusson S, Hardardottir H, Harvey RP, Lai D, Zhou M, Brunner D, Mutel V, Gonzalo A, Lemke G, Sainz J, Johannesson G, Andresson T, Gudbjartsson D, Manolescu A, Frigge ML, Gurney ME, Kong A, Gulcher JR, Petursson H, Stefansson K. Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet*. 2002;71(4):877-892.
- Stefansson H, Sarginson J, Kong A, Yates P, Steinthorsdottir V, Gudfinnsson E, Gunnarsdottir S, Walker N, Petursson H, Crombie C, Ingason A, Gulcher JR, Stefansson K, St Clair D. Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. *Am J Hum Genet*. 2003;72(1):83-87.
- Yang JZ, Si TM, Ruan Y, Ling YS, Han YH, Wang XL, Zhou M, Zhang HY, Kong QM, Liu C, Zhang DR, Yu YQ, Liu SZ, Ju GZ, Shu L, Ma DL, Zhang D. Association study of neuregulin 1 gene with schizophrenia. *Mol Psychiatry*. 2003;8(7):706-709.
- Corvin AP, Morris DW, McGhee K, Schwaiger S, Scully P, Quinn J, Meagher D, Clair DS, Waddington JL, Gill M. Confirmation and refinement of an 'at-risk' haplotype for schizophrenia suggests the EST cluster, Hs.97362, as a potential susceptibility gene at the Neuregulin-1 locus. *Mol Psychiatry*. 2004;9(2):208-213.
- Li T, Stefansson H, Gudfinnsson E, Cai G, Liu X, Murray RM, Steinthorsdottir V, Januel D, Gudnadottir VG, Petursson H, Ingason A, Gulcher JR, Stefansson K, Collier DA. Identification of a novel neuregulin 1 at-risk haplotype in Han schizophrenia Chinese patients, but no association with the Icelandic/Scottish risk haplotype. *Mol Psychiatry*. 2004;9(7):698-704.
- Hall D, Gogos JA, Karayiorgou M. The contribution of three strong candidate schizophrenia susceptibility genes in demographically distinct populations. *Genes Brain Behav*. 2004;3(4):240-248.
- Petryshen TL, Middleton FA, Kirby A, Aldinger KA, Purcell S, Tahl AR, Morley CP, McGann L, Gentile KL, Rockwell GN, Medeiros HM, Carvalho C, Macedo A, Dourado A, Valente J, Ferreira CP, Patterson NJ, Azevedo MH, Daly MJ, Pato CN, Pato MT, Sklar P. Support for involvement of neuregulin 1 in schizophrenia pathophysiology. *Mol Psychiatry*. 2005;10(4):366-374.
- Lachman HM, Pedrosa E, Nolan KA, Glass M, Ye K, Saito T. Analysis of polymorphisms in AT-rich domains of neuregulin 1 gene in schizophrenia. *Am J Med Genet B Neuropsychiatr Genet*. 2006;141B(1):102-109.
- Fukui N, Muratake T, Kaneko N, Amagane H, Someya T. Supportive evidence for neuregulin 1 as a susceptibility gene for schizophrenia in a Japanese population. *Neurosci Lett*. 2006;396(2):117-120.
- Kim JW, Lee YS, Cho EY, Jang YL, Park DY, Choi KS, Jeun HO, Cho SH, Jang SY, Hong KS. Linkage and association of schizophrenia with genetic variations in the locus of neuregulin 1 in Korean population. *Am J Med Genet B Neuropsychiatr Genet*. 2006;141B(3):281-286.
- Thomson PA, Christoforou A, Morris SW, Adie E, Pickard BS, Porteous DJ, Muir WJ, Blackwood DH, Evans KL. Association of neuregulin 1 with schizophrenia and bipolar disorder in a second cohort from the Scottish population. *Mol Psychiatry*. 2007;12(1):94-104.
- Turunen JA, Peltonen JO, Pietiläinen OP, Henna W, Loukola A, Paunio T, Sillander K, Ekelund J, Varilo T, Partonen T, Lönnqvist J, Peltonen L. The role of DTNBP1, NRG1, and AKT1 in the genetics of schizophrenia in Finland. *Schizophr Res*. 2007;91(1-3):27-36.
- Thiselton DL, Webb BT, Neale BM, Ribble RC, O'Neill FA, Walsh D, Riley BP, Kendler KS. No evidence for linkage or association of neuregulin-1 (*NRG1*) with disease in the Irish study of high-density schizophrenia families (ISHDSF). *Mol Psychiatry*. 2004;9(8):777-783.
- Hong CJ, Huo SJ, Liao DL, Lee K, Wu JY, Tsai SJ. Case-control and family-based association studies between the neuregulin 1 (Arg38Gln) polymorphism and schizophrenia. *Neurosci Lett*. 2004;366(2):158-161.
- Duan J, Martinez M, Sanders AR, Hou C, Krasner AJ, Schwartz DB, Gejman PV.

- Neuregulin 1 (NRG1) and schizophrenia: analysis of a US family sample and the evidence in the balance. *Psychol Med*. 2005;35(11):1599-1610.
16. Ingason A, S  by K, Timm S, Wang AG, Jakobsen KD, Fink-Jensen A, Hemmingsen R, Berg Rasmussen H, Werge T. No significant association of the 5' end of neuregulin 1 and schizophrenia in a large Danish sample. *Schizophr Res*. 2006;83(1):1-5.
  17. Walss-Bass C, Raventos H, Montero AP, Armas R, Dassori A, Contreras S, Liu W, Medina R, Levinson DF, Pereira M, Leach RJ, Almasy L, Escamilla MA. Association analyses of the neuregulin 1 gene with schizophrenia and manic psychosis in a Hispanic population. *Acta Psychiatr Scand*. 2006;113(4):314-321.
  18. Vilella E, Costas J, Sanjuan J, Guitart M, De Diego Y, Carracedo A, Martorell L, Valero J, Labad A, De Frutos R, N  jera C, Molt   MD, Toirac I, Guillaumat R, Brunet A, Vall  s V, P  rez L, Leon M, de Fonseca FR, Phillips C, Torres M. Association of schizophrenia with DTNBP1 but not with DAO, DAOA, NRG1 and RGS4 nor their genetic interaction. *J Psychiatr Res*. 2008;42(4):278-288.
  19. Rosa A, Gardner M, Cuesta MJ, Peralta V, Fatj  -Vilas M, Miret S, Navarro ME, Comas D, Fa  n  s L. Family-based association study of neuregulin-1 gene and psychosis in a Spanish sample. *Am J Med Genet B Neuropsychiatr Genet*. 2007;144B(7):954-957.
  20. Sanders AR, Duan J, Levinson DF, Shi J, He D, Hou C, Burrell GJ, Rice JP, Nertney DA, Olincy A, Rozic P, Vinogradov S, Buccola NG, Mowry BJ, Freedman R, Amin F, Black DW, Silverman JM, Byerley WF, Crowe RR, Cloninger CR, Martinez M, Gejman PV. No significant association of 14 candidate genes with schizophrenia in a large European ancestry sample: implications for psychiatric genetics. *Am J Psychiatry*. 2008;165(4):497-506.
  21. 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(12):1995-2002.
  22. Munaf   MR, Attwood AS, Flint J. Neuregulin 1 genotype and schizophrenia. *Schizophr Bull*. 2008;34(1):9-12.
  23. Norton N, Moskvina V, Morris DW, Bray NJ, Zammit S, Williams NM, Williams JJ, Preece AC, Dwyer S, Wilkinson JC, Spurlock G, Kirov G, Buckland P, Waddington JL, Gill M, Corvin AP, Owen MJ, O'Donovan MC. Evidence that interaction between neuregulin 1 and its receptor erbB4 increases susceptibility to schizophrenia. *Am J Med Genet B Neuropsychiatr Genet*. 2006;141B(1):96-101.
  24. Shiota S, Tochigi M, Shimada H, Ohashi J, Kasai K, Kato N, Tokunaga K, Sasaki T. Association and interaction analyses of NRG1 and ERBB4 genes with schizophrenia in a Japanese population. *J Hum Genet*. 2008;53(10):929-935.
  25. Benzel I, Bansal A, Browning BL, Galwey NW, Maycox PR, McGinnis R, Smart D, St Clair D, Yates P, Purvis I. Interactions among genes in the ErbB-Neuregulin signalling network are associated with increased susceptibility to schizophrenia. *Behav Brain Funct*. 2007;3:31.
  26. Shao H, Burrage LC, Sinasac DS, Hill AE, Ernest SR, O'Brien W, Courtland HW, Jepsen KJ, Kirby A, Kulbokas EJ, Daly MJ, Broman KW, Lander ES, Nadeau JH. Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proc Natl Acad Sci U S A*. 2008;105(50):19910-19914.
  27. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/A:1010933404324.
  28. Hothorn T, Lausen B, Benner A, Radespiel-Tr  ger M. Bagging survival trees. *Stat Med*. 2004;23(1):77-91.
  29. Hothorn T, B  hlmann P, Dudoit S, Molinaro A, van der Laan MJ. Survival ensembles. *Bioinformatics*. 2006;7(3):355-373.
  30. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*. 2007;8:25.
  31. Kooperberg C, Ruczinski I. Identifying interacting SNPs using Monte Carlo logic regression. *Genet Epidemiol*. 2005;28(2):157-170.
  32. Nicodemus KK, Callicott JH, Higier RG, Luna A, Nixon DC, Lipska BK, Vakkalanka R, Giegling I, Rujescu D, St. Clair D, Muglia P, Shugart YY, Weinberger DR. Evidence of statistical epistasis between DISC1, CIT and NDEL1 impacting risk for schizophrenia: biological validation with functional neuroimaging. *Hum Genet*. 2010;127(4):441-452.
  33. Meyer-Lindenberg A, Weinberger DR. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nat Rev Neurosci*. 2006;7(10):818-827.
  34. Hahn CG, Wang HY, Cho DS, Talbot K, Gur RE, Berrettini WH, Bakshi K, Kamins J, Borgmann-Winter KE, Siegel SJ, Gallop RJ, Arnold SE. Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor hypofunction in schizophrenia. *Nat Med*. 2006;12(7):824-828.
  35. Mei L, Xiong WC. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. *Nat Rev Neurosci*. 2008;9(6):437-452.
  36. Li B, Woo RS, Mei L, Malinow R. The neuregulin-1 receptor erbB4 controls glutamatergic synapse maturation and plasticity. *Neuron*. 2007;54(4):583-597.
  37. Nagata K, Kohda D, Hatanaka H, Ichikawa S, Matsuda S, Yamamoto T, Suzuki A, Inagaki F. Solution structure of the epidermal growth factor-like domain of heregulin-alpha, a ligand for p180erbB-4. *EMBO J*. 1994;13(15):3517-3523.
  38. Bao J, Lin H, Ouyang Y, Lei D, Osman A, Kim TW, Mei L, Dai P, Ohlemiller KK, Ambron RT. Activity-dependent transcription regulation of PSD-95 by neuregulin-1 and Eos. *Nat Neurosci*. 2004;7(11):1250-1258.
  39. Brennan JE, Chao DS, Gee SH, McGee AW, Craven SE, Santillano DR, Wu Z, Huang F, Xia H, Peters MF, Froehner SC, Bredt DS. Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains. *Cell*. 1996;84(5):757-767.
  40. Jaffrey SR, Snowman AM, Eliasson MJ, Cohen NA, Snyder SH. CAPON: a protein associated with neuronal nitric oxide synthase that regulates its interactions with PSD95. *Neuron*. 1998;20(1):115-124.
  41. Brzustowicz LM, Simone J, Mohseni P, Hayter JE, Hodgkinson KA, Chow EW, Bassett AS. Linkage disequilibrium mapping of schizophrenia susceptibility to the CAPON region of chromosome 1q22. *Am J Hum Genet*. 2004;74(5):1057-1063.
  42. Shinkai T, Ohmori O, Hori H, Nakamura J. Allelic association of the neuronal nitric oxide synthase (NOS1) gene with schizophrenia. *Mol Psychiatry*. 2002;7(6):560-563.
  43. Xu MQ, Xing QH, Zheng YL, Li S, Gao JJ, He G, Guo TW, Feng GY, Xu F, He L. Association of AKT1 gene polymorphisms with risk of schizophrenia and with response to antipsychotics in the Chinese population. *J Clin Psychiatry*. 2007;68(9):1358-1367.
  44. Egan MF, Goldberg TE, Gscheidle T, Weirich M, Bigelow LB, Weinberger DR. Relative risk of attention deficits in siblings of patients with schizophrenia. *Am J Psychiatry*. 2000;157(8):1309-1316.
  45. Need AC, Ge D, Weale ME, Maia J, Feng S, Heinzen EL, Shianna KV, Yoon W, Kasperavi  ute D, Gennarelli M, Strittmatter WJ, Bonvicini C, Rossi G, Jayatilake K, Cola PA, McEvoy JP, Keefe RS, Fisher EM, St Jean PL, Giegling I, Hartmann AM, M  ller HJ, Ruppert A, Fraser G, Crombie C, Middleton LT, St Clair D, Roses AD, Muglia P, Francks C, Rujescu D, Meltzer HY, Goldstein DB. A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS Genet*. 2009;5(2):e1000373.
  46. Nicodemus KK, Wang W, Shugart YY. Stability of variable importance scores and rankings using statistical learning tools on single-nucleotide polymorphisms and risk factors involved in gene x gene and gene x environment interactions. *BMC Proc*. 2007;1(suppl 1):S58.
  47. Nicodemus KK, Malley JD. Predictor correlation impacts machine learning algorithms: implications for genomic studies. *Bioinformatics*. 2009;25(15):1884-1890.
  48. Callicott JH, Mattay VS, Bertolino A, Finn K, Coppola R, Frank JA, Goldberg TE, Weinberger DR. Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cereb Cortex*. 1999;9(1):20-26.
  49. Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp*. 1996;4(1):58-73.
  50. Fardo DW, Becker KD, Bertram L, Tanzi RE, Lange C. Recovering unused information in genome-wide association studies: the benefit of analyzing SNPs out of Hardy-Weinberg equilibrium. *Eur J Hum Genet*. 2009;17(12):1676-1682.
  51. Blokland GA, McMahon KL, Hoffman J, Zhu G, Meredith M, Martin NG, Thompson PM, de Zubicaray GI, Wright MJ. Quantifying the heritability of task-related brain activation and performance during the N-back working memory task: a twin fMRI study. *Biol Psychol*. 2008;79(1):70-79.
  52. Suarez BK, Duan J, Sanders AR, Hinrichs AL, Jin CH, Hou C, Buccola NG, Hale N, Weillbaecher AN, Nertney DA, Olincy A, Green S, Schaffer AW, Smith CJ, Hannah DE, Rice JP, Cox NJ, Martinez M, Mowry BJ, Amin F, Silverman JM, Black DW, Byerley WF, Crowe RR, Freedman R, Cloninger CR, Levinson DF, Gejman PV. Genomewide linkage scan of 409 European-ancestry and African American families with schizophrenia: suggestive evidence of linkage at 8p23.3-p21.2 and 11p13.1-q14.1 in the combined sample. *Am J Hum Genet*. 2006;78(2):315-333.
  53. Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ, Tanzi RE, Bertram L. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet*. 2008;40(7):827-834.
  54. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet*. 2009;10(6):392-404.
  55. Meyer-Lindenberg A, Nicodemus KK, Egan MF, Callicott JH, Mattay V, Weinberger DR. False positives in imaging genetics. *Neuroimage*. 2008;40(2):655-661.
  56. Tan HY, Callicott JH, Weinberger DR. Intermediate phenotypes in schizophrenia genetics redux: is it a no brainer? *Mol Psychiatry*. 2008;13(3):233-238.
  57. Weinberger DR, Egan MF, Bertolino A, Callicott JH, Mattay VS, Lipska BK, Berman KF, Goldberg TE. Prefrontal neurons and the genetics of schizophrenia. *Biol Psychiatry*. 2001;50(11):825-844.
  58. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P; International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748-752.