# Deep resequencing and association analysis of schizophrenia candidate genes

Christopher E Hilliard[1], James J Crowley[1], Yunjung Kim[1], Margaret B Morgan[2], Jeffrey A Lieberman[3], Patrick F Sullivan[1], Richard A Gibbs[2]

*[1]Departments of Genetics, University of North Carolina, Chapel Hill, NC; [2]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; [3]Department of Psychiatry, Columbia University, New York, NY*

## Abstract

**Introduction:** In 2005, we selected 10 genes for which there was reasonable evidence for involvement in the etiology of schizophrenia (*COMT*, *DAOA*, *DISC1*, *DRD2*, *DRD3*, *DTNBP1*, *HTR2A*, *NRG1*, *SLC6A3*, *SLC6A4*, PMID 16033310). Although these genes have not received support from subsequent GWAS and may not contain etiological common variation, it is possible that they contain uncommon variation of etiological importance.
**Methodology:** We conducted a multistage study. In Stage 1, we used Sanger methods to sequence the exons, 5' and 3' UTRs, splice sites, promoters and conserved intronic regions of these 10 genes in 727 cases with schizophrenia from CATIE and 733 controls (EUR and AFR ancestry). In Stage 2, we validated these SNPs using Roche 454 sequencing in the same samples. In Stage 3, we used Sequenom iPlex to genotype prioritized SNPs in independent samples, 2,192 schizophrenia cases and 2,659 controls from the Molecular Genetics of Schizophrenia Collaboration.
**Results:** Sanger sequencing (Stage 1) identified 753 SNPs (including 558 novel variants). We prioritized 92 SNPs for Stage 3, and 77 were successfully genotyped. No SNP showed association with schizophrenia either individually or using methods to aggregate exonic variation within genes.
**Conclusions:** We found no association evidence to support the involvement of these genes in schizophrenia.

## Stage 1: Discovery

- This work was part of the NHGRI Medical Resequencing Project. In 2005, we selected 10 genes with some evidence of involvement in the etiology of schizophrenia (SCZ) for sequencing (Table 1). Figure 1 shows the overall strategy.
- Sanger dideoxy sequencing of exons, 5′ and 3′ UTRs, splice sites, promoters and conserved intronic regions was done at Baylor to discover novel variants.
- Discovery sample: DNA from LBCLs from 727 SCZ cases from CATIE and 733 group-matched controls. Cases: 74% male, ancestry 57% EUR, 29% AFR, and 14% other ancestry. Controls: 67% males, with similar ancestry as cases.
- Individual sequencing of 99 amplicons encompassing 16,952 bp. SNPDetector was used to identify SNPs and insertions/deletions.
- 753 SNPs discovered; 73% not in dbSNP132, and 30% predicted to be functional. 169 variants were missense, and 3 were nonsense.

| Gene | Product | # amplicons | kb sequenced |
|---|---|---|---|
| *COMT* | catechol-O-methyltransferase | 4 | 0.8 |
| *DAOA* | D-amino acid oxidase activator | 10 | 0.7 |
| *DISC1* | disrupted in schizophrenia 1 | 15 | 2.7 |
| *DRD2* | dopamine receptor D2 | 7 | 1.4 |
| *DRD3* | dopamine receptor D3 | 6 | 1.2 |
| *DTNBP1* | dystrobrevin binding protein 1 | 10 | 1.2 |
| *HTR2A* | serotonin receptor 2A | 3 | 1.4 |
| *NRG1* | neuregulin 1 | 17 | 3.6 |
| *SLC6A3* | dopamine transporter | 14 | 1.9 |
| *SLC6A4* | serotonin transporter | 13 | 1.9 |
| **Totals:** | **10 genes** | **99** | **16.8** |

**Table 1.** 10 genes showing evidence for involvement in etiology of SCZ as of 2005. The regions of interest contain intronic, synonymous and missense variants, as well as mutations in the 5' and 3' UTR, proximal promoter, and splice region.

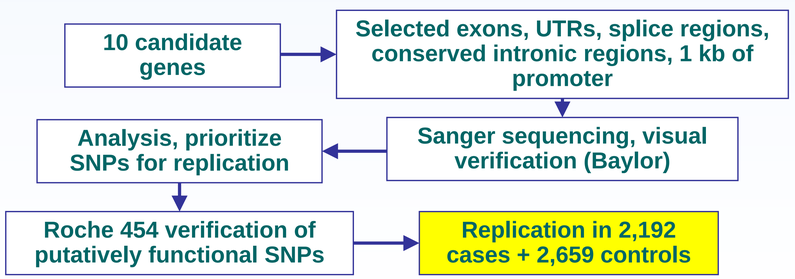


**Figure 1.** Overall strategy.

## Stage 2: Validation

- 254 of 753 SNPs chosen for technical replication using Roche 454 sequencing
- 254 SNPs were: (1) novel nonsense, missense or splice site variant, (2) in dbSNP but predicted nonsense, missense or splice site variant with 2X (forward and reverse strand) coverage in ≥1 case of EUR ancestry, (3) novel intronic variant with 2X coverage in a case of EUR ancestry, or (4) novel variant with an odds ratio >2.
- PCR products from 254 samples representing all 254 minor alleles were pooled and sequenced. A variant was validated if the minor allele was detected in this pool.
- Roche 454 sequencing validated 225/254 variants (accuracy 89%).

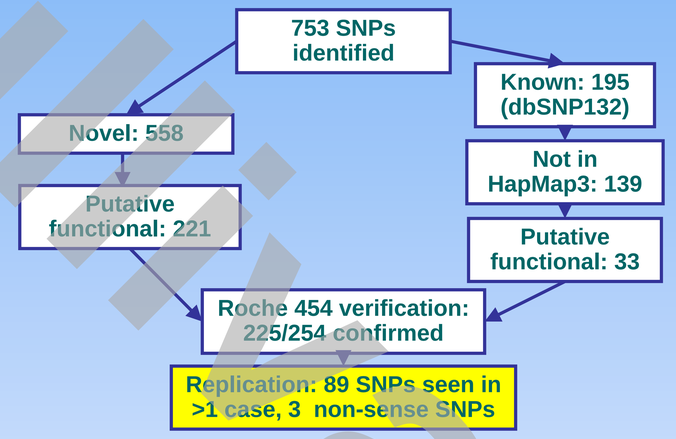


**Figure 2.** SNP workflow.

## Stage 3: Replication

- Used Sequenom iPlex to genotype prioritized SNPs in independent samples.
- Cases. EUR and AFR unrelated adult cases with DSM-IV SCZ or schizoaffective, part of the Molecular Genetics of Schizophrenia Part 2 (MGS2, PMID 18198266). 2,192 cases were genotyped in this study. Cases were 71% EUR and 29% AFR. Cases were 68% male. DNA from LCBLs.
- Controls. EUR and AFR unrelated adult controls collected by Knowledge Networks (Menlo Park, CA). Briefly, KN uses random digit dialing to area codes selected to represent the national population, to recruit individuals to join a nationwide survey panel. Subjects were excluded if they had a lifetime diagnosis or treatment of SCZ or schizoaffective disorder, or hallucinations or delusions, or of bipolar disorder. 2,659 matched controls were genotyped in this study. Controls were 71% EUR and 29% AFR. Controls were 43% male. DNA from LCBLs.
- SNPs selected: (1) validated in Stage 2, (2) observed in ≥1 Stage 1 case, or (3) predicted nonsense variant 57 missense, 13 intronic, 6 synonomous, 4 in the 5' UTR, 3 in the 3' UTR, 3 nonsense, 3 proximal promoter and 3 splice region variants.
- Genotyping was with Sequenom iPlex MassARRAY, three plexes were designed using SpectroDESIGNER and genotypes were called using TyperANALYZER. Technical validity: 92 SNPs were re-genotyped in the Stage 1 samples, and concordance with Sanger sequencing was 99.7%

## Acknowledgements

Funding: NGGRI Medical Resequencing Program

- Association tests of 77 SNPs (dropped 15, monomorphic or poor performance)
- Single-marker analysis for 77 SNPs was performed separately by ancestry using logistic regression in PLINK
- Aggregate effect of uncommon variation tested using PLINK/SEQ, C-ALPHA (PMID 21408211) and Variable-Threshold (PMID 20471002) analyses for 45 non-intronic SNPs with MAF <0.01.

| Gene Name | Single-marker analysis (77 SNPs) | | | Gene-based analysis (45 SNPs) | | | | |
|---|---|---|---|---|---|---|---|---|
| | # of SNPs | EUR (Minimum p-value) | AFR (Minimum p-value) | # of uncommon variants | EUR (# of case/control alternative allele) | EUR p-values (CALPHA/VT) | AFR (# of case/control alternative allele) | AFR p-values (CALPHA/VT) |
| ***COMT*** | 5 | 0.798 | 0.299 | 3 | 8/1 | 0.02/0.003 | 0/3 | 0.29/0.71 |
| ***DAOA*** | 5 | 0.916 | 2.7E-6 | 0 | - | - | - | - |
| ***DISC1*** | 34 | 0.368 | 0.381 | 30 | 32/40 | 0.67/0.33 | 39/32 | 0.27/0.08 |
| ***DRD3*** | 3 | 0.375 | 0.739 | 0 | - | - | - | - |
| ***DTNBP1*** | 11 | 0.390 | 0.115 | 5 | 18/19 | 0.32/0.72 | 7/4 | 0.34/0.16 |
| ***HTR2A*** | 3 | 0.448 | 0.739 | 1 | 3/13 | 0.03/0.7 | 0/1 | 0.71/0.71 |
| ***NRG1*** | 8 | 0.127 | 0.017 | 1 | 4/1 | 0.14/0.1 | 159/185 | 0.34/0.15 |
| ***SLC6A3*** | 4 | - | - | 2 | 4/4 | 0.6/0.8 | 0/1 | 0.76/0.67 |
| ***SLC6A4*** | 4 | 0.885 | 0.677 | 3 | 10/9 | 0.45/0.25 | 1/1 | 0.4/0.36 |

**Table 2.** Single-marker and gene-based analyses on EUR and AFR samples. Logistic regression performed on 77 SNPs. Gene-based analysis performed on 45 SNPs in PLINK/SEQ with C-ALPHA and VT tests.

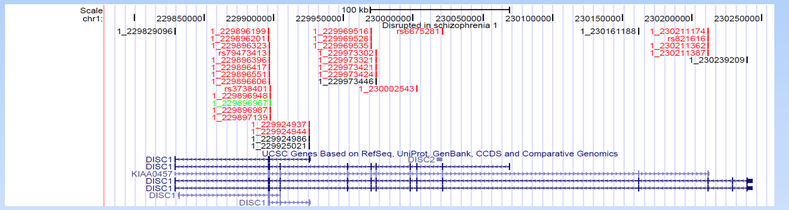


**Figure 3.** *DISC1* variants genotyped in replication sample (hg18). Red=misssense/nonsense, green=synonymous, black=splice site.

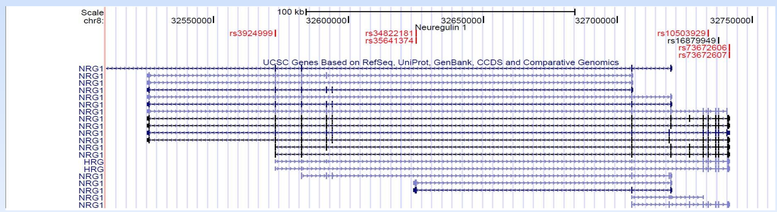


**Figure 4.** *NRG1* variants genotyped in replication sample (hg18). Red=misssense/nonsense, green=synonymous, black=splice site.

## Conclusions

- Classical SCZ candidate genes contain variation not directly genotyped on current GWAS chips.
- There was some initial evidence *DISC1* had more novel SNPs in cases (presented WCPG San Diego), but this did not replicate. *DISC1* was not found to contain common or uncommon variants individually or in aggregate associated with SCZ.
- Despite testing 77 suggestive variants in a replication sample 3 times larger than the discovery sample, no compelling associations beyond chance were discovered.
- We found no evidence that uncommon or common genetic variation in these 10 genes plays a role in SCZ etiology.

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL