# Genome-Wide Association Study of Schizophrenia

Pablo V. Gejman M.D

ENH / Northwestern University
Evanston, IL

# Overall description

## Genomewide association study of schizophrenia (SZ)

- **GAIN sample:**
  - 1450 / 1450 European Ancestry (EA) cases / controls.
    (Affymetrix Mapping 500K Array)
  - 1200 / 1200 African American (AA) cases / controls.
    (Affymetrix Mapping 1M Array)
- **Additional available subjects:**
  - 1550 / 1550 EA cases / controls.
- **Total sample of 3000 / 3000 EA and 1200 / 1200 AA**
- **DNA and clinical data for all cases and controls are or will be publicly available from the NIMH repository program (http://www.nimhgenetics.org).**

# Conceptualization of SZ phenotype and evidence for genetic influence

- Considerable data suggest SZ is a neurodevelopmental disorder.
- Onset typically in teens or early adulthood, 0.5-1% prevalence.
- High heritability (80-85%); high MZ:DZ concordance ratio (~0.5:0.1) suggesting multigenic transmission.
- Possible environmental factors (perinatal hypoxia; intra-uterine viral or immune effects; migration), none definitively proven.
- Family studies: relative risk (RR) to first-degree relatives is ≈ 10.
- Family studies: co-aggregation of SZ, schizoaffective, other non-affective psychoses, and schizotypal, schizoid and paranoid personality disorders.
- Adoption studies: risk travels with biological relationship

# Characteristics of SZ relevant to the definition of the primary affection status

- **Symptoms:** hallucinations and/or delusions, disorganized thinking / speech and behavior, loss of emotions and interest.
- **Course:** Generally chronic. Medications seldom produce full remissions.
- **Inclusion of SZ and SA cases:**
  - Almost half of SZ patients report depressive and/or manic periods.
  - Current criteria define schizoaffective (SA) disorder as co-occurring SZ plus depressive and/or manic syndromes for 'substantial' periods, but SZ persists.
  - The criterion for relative duration of mood disorder / psychosis is subjective in the DSM classification, and difficult to apply reliably (requiring an arbitrary cut-off).
  - SZ and SA co-segregate in families. Both respond to anti-psychotic medications. Trend toward over-diagnosis of SA (perhaps to justify use of medications for mood disorder).
  - The combination of SZ/SA has very high diagnostic reliability

# Cases

- **Sources:** treatment setting (outpatient, residential, inpatient) and family advocacy/support groups (NAMI).
- **Psychiatric assessment:** Diagnostic Interview for Genetic Studies (DIGS) and Medical Records (over 99%); informant interview where possible (FIGS).
- **Consensus diagnosis of SZ or SA:** after independent review by 2 diagnosticians, ≈90% are confirmed cases.
- **Cases with DNA:** 3,009 EA and 1,383 AA to date.
- **Repository clinical data:** demographics (no identifiers); DIGS ratings; dimensional scales; family history of psychosis, SZ and mania; comorbid psychiatric substance use diagnoses; age of onset; age of parents at birth
- **Representativeness:** Our sample is likely to be representative of SZ in the community: chronic unremitting cases, both SZ and SA diagnoses with similar ages at onset and SZ symptomatology

## Results of SZ linkage Genome Scan Meta-Analysis Show Trend Towards Convergence of Linkage Signals (GSMA, Levinson et al., 2003; Lewis et al., 2003)

12 30-cM bins in 10 regions met empirical criteria for genomewide significance in ≈ 1200 families.

7 additional regions were suggestive by one of two tests.

3 regions → suggestive linkage in our MGS1 scan.

| Cytogenetic location | Bin | Rank (weighted) | By-bin | Ordered-rank | Both <0.05 | MGS1 scan |
|---|---|---|---|---|---|---|
| 2p12-q22.1 | 2.5 | 1 | *0.0004* | *0.0327* | ▲ | |
| 5q23.2-q34 | 5.5 | 2 | *0.0032* | *0.0491* | ▲ | |
| 3p25.3-p22.1 | 3.2 | 3 | *0.0060* | *0.0311* | ▲ | |
| 11q22.3-q24.1 | 11.5 | 4 | *0.0060* | *0.0040* | ▲ | ++ |
| 6pter-p22.3 | 6.1 | 5 | *0.0159* | *0.0328* | ▲ | |
| 2q22.1-q23.3 | 2.6 | 6 | *0.0230* | *0.0448* | ▲ | |
| 1p13.3-q23.3 | 1.6 | 7 | *0.0235* | *0.0136* | ▲ | |
| 22pter-q12.3 | 22.1 | 8 | *0.0310* | *0.0216* | ▲ | |
| 8p22-p21.1 | 8.2 | 9 | *0.0310* | *0.0068* | ▲ | ++ |
| 6p22.3-p21.1 | 6.2 | 10 | *0.0330* | *0.0024* | ▲ | |
| 20p12.3-p11 | 20.2 | 11 | *0.0460* | *0.0098* | ▲ | |
| 14pter-q13.1 | 14.1 | 12 | *0.0470* | *0.0043* | ▲ | |
| 16p13-q12.2 | 16.2 | 13 | 0.0560 | *0.0069* | | |
| 18q22.1-qter | 18.4 | 14 | 0.0650 | *0.0103* | | |
| 10pter-p14 | 10.1 | 15 | 0.0680 | *0.0046* | | + |
| 1q23.3-q31.1 | 1.7 | 16 | 0.0820 | *0.0142* | | |
| 15q21.3-q26.1 | 15.3 | 17 | 0.0950 | *0.0293* | | |
| 6q15-q23.2 | 6.4 | 18 | 0.0980 | *0.0177* | | |
| 17q21.33-q24.3 | 17.3 | 19 | 0.1120 | *0.0349* | | |

- Our MGS1 study (Suarez et al., 2004) of ~ 500 ASPs (not included in GSMA): STRP scan, SNP fine-mapping → largest peak on 8p23.3-p21.2 (EA+AA and EA alone)
- The Multicenter Genetic Studies of SZ study (Levinson et al.), partially overlapping with the above GSMA, recently reported its largest peak on 8p.
- GSMA is now being updated to add 10 new studies (≈1400 families)

# SZ candidate genes identified in regions of linkage or cytogenetic abnormalities:

| Gene (location) | Discovery sample | Discovery strategy | GSMA region |
|---|---|---|---|
| DTNBP1 (6p22.3) | Irish families | Linkage/LD mapping | Yes |
| NRG1 (8p12) | Icelandic families | Linkage/LD mapping | Yes |
| DAOA (13q33.2-q34) | US, Canadian families; 2 case-control samples | Linkage/LD mapping | No |
| TAAR6/STX7 6q23.2) | U.S. families | Linkage/LD mapping | Yes |
| CAPON (1q22) | Canadian families | Linkage/LD mapping | Yes |
| CHRNA7 (15q14) | U.S. families | Linkage/cand gene/endophenotype | No |
| COMT/ARVCF (22q11.21) | Multiple association studies | Deletion (VCFS) | Yes |
| DISC1 (1q41.2) | Large Scottish pedigree | Translocation | No |

- Reported associations implicate common alleles.
- Replication has been inconsistent.
- “Rare allele” hypotheses not systematically tested.
- Mechanistic candidates: Associations (ORs 1.15-1.3) reported in meta-analyses (such as functional polymorphisms in DRD2 --Ser311Cys).

## Molecular Genetics of SZ (MGS) 1 and 2: Results

- **MGS1: Linkage scan of 263 EA, 146 AA families (503 ASPs)**
  - 9 cM STRP scan by CIDR, dense LD mapping with SNPs by ENH/Northwestern
  - Suggestive linkage: 8p and 11cen (EA+AA); 4p and 5p (EA only)
  - Suarez et al., AJHG 2006;78:315-333
- **MGS2 8p experiment (genotyping completed, analyses in progress):**
  - LD mapping of 28 Mb candidate region on 8p23.3-p21.2 at Illumina
  - 1,500 SNPs (Illumina) in 1650 EA cases / 2100 controls; 160 genes
- **MGS2 candidate gene experiment (same sample, SNPlex/Taqman):**
  - 14 candidate genes (RGS4, AKT1, COMT, DAOA, ARVCF, TRAR4/STX7, HTR2A, DRD2, PPP3CC, CHRNA7, DTNBP1, NRG1, DISC1 --DISC1 & NRG1 analysis in progress), 2.27 MB, 843 SNPs (cnsSNPs, nsSNPs, additional SNPs in or near every exon and regulatory sequence), plus 215 ancestry informative markers; AIMs). Genotyping: ENH/Northwestern
  - Primary analysis: single SNPs, EIGENSTRAT. No significant findings (corrected).

# Controls: recruitment and screening

- **Source:** Representative national EA and AA samples recruited by survey research company (Knowledge Networks, Inc.)
- **Online** consent and self-report clinical questionnaire (subjects with questionnaire responses suggesting a psychiatric disorder remain in the control dataset at the discretion of the user).
- **Phlebotomy** performed at home; additional in-person consent.
- **Accrued sample:** 3,364 EA and 1,190 AA controls

  - AA recruitment is ongoing, target = 1,320 by the end of 2006
- **8% excluded from our study** for endorsement of lifetime SZ/SA or psychosis, bipolar disorder, or failure to answer any of these questions

# Controls: Clinical questionnaire data

- **CIDI-Short Form** (Composite Instrument for Diagnostic Interviewing, World Health Organization):
  - Major depression, generalized anxiety, specific phobia, panic attack, agoraphobia and social phobia, alcohol dependence and drug dependence
- Single items about lifetime diagnosis of (or treatment for) **psychotic or bipolar disorders**
- Others: one question about **sexual orientation;** 12-item Eysenck **neuroticism and extraversion** (personality) scales; **height and weight**, and highest lifetime weight outside pregnancy; brief **nicotine** dependence inventory

# Basis for matching cases and controls

- ***Primary matching variable:*** EA or AA ancestry
- ***Why have we not ascertained in a systematic manner?***
  - Funded to collect 4,500 SZ cases and 4,500 controls, largest sample ever attempted. Prevalence too low for systematic ascertainment. Needed high ratio of ascertained subjects / eligible cases
- ***Why are there no data on environmental exposures?***
  - Standard lifestyle variables have no known effects
  - Relevant variables are difficult to measure retrospectively (perinatal events; intra-uterine exposures); expensive to measure (brain imaging tests for neurodevelopmental changes); or would require stratification into less powerful subsamples (recent migration).
- ***Are recruited cases similar to untreated and non-cooperative cases?***
  - Yes: most subjects had past periods when we could not have recruited them due to clinical severity, homelessness, incarceration, etc.
- ***Are controls representative of the population?***
  - Demographically representative except for small excess of higher education and SES.

## *Why did we not match for:*

- ***SES?***
  - SZ causes occupational and social impairments resulting in lower SES.
  - Information on parental SES would be unreliable (no family informant for many cases).
  - Stratifying by cases' SES would select unusual controls (cognitive and behavioral traits).
- ***Sex?***
  - Same SZ prevalence in males and females; no established hormonal. abnormalities; onset not correlated with puberty or menopause.
  - Preponderance of males is related to recruitment from clinical caseloads (males have earlier age of onset and worse course).
- ***Age?***
  - Matching by cases' age would select more controls who were still in age of risk.
  - Older controls could create false positive case-control differences, but this is unlikely for most SNPs.

- NIMH repository controls were also intended for use in studies of multiple psychiatric disorders, so emphasis was on population representativeness.

# Restrictions on data use

- **Consent forms for cases and controls specifically authorize:**
  - Depositing of the blood specimen in the NIMH DNA and Cell Repository,
  - Creation of cell lines and extraction of DNA,
  - Depositing of deidentified epidemiological and clinical information
  - Sharing of biological materials and clinical data with qualified scientists
- **Controls are fully anonymized 1 year after collection**
- **Restrictions:**
  - GAIN regulations for accessing data
  - NIMH restrictions for accessing DNA and epidemiological and clinical data from the NIMH repository

# Collaborators

- **Lead Investigator** -- Gejman PV[1]
- **Statistical Task Force Chair** --Levinson DF[2]
- **Clinical Database** -- Levinson DF[2]; Sanders AR[1] (**Lead PI Assistant, Curator**)
- **Laboratory, Molecular Database** -- Duan J[1], Gejman PV[1]
- **Statisticians** -- Dudbridge F[3], Martinez M[4], Holmans PA[5], Whittemore AS[2], Rice JP[6], Pe'er I[7], Jianxin Shi[2], Jun Li[2]
- **Clinical Site PI's** -- Mowry BJ[8], Buccola NG[9], Amin F[10], Black DW[11], Silverman JM[12], Byerley WF[13], Cloninger CR[6], Freedman R[14]·Levinson DF[2], Gejman PV[1]
- **Collaborator, Clinical Analyses** -- Kendler KS[15], Fanous AH[16]
- **Collaborators, populations** --Shriver M[17], Mc.Keigue P[18], Kidd K[19]

1- ENH & Northwestern U, Evanston, IL
2- Stanford U, Palo Alto, CA
3- MRC Biostatistics Unit, London, England
4- INSERM, Toulouse, France
5- Wales College Med./Cardiff U., Cardiff, Wales
6- Washington U, St. Louis, MO
7- Columbia University
8- QCSR and U Queensland, Brisbane, Aus
9- LSU Health Sciences Ctr, New Orleans, LA
10- Atlanta VA Med Ctr & Emory U, Atlanta, GA
11- Univ. Iowa, Iowa City, IA
12- Mt. Sinai School of Medicine, New York, NY
13- UCSF, San Francisco, CA
14- U Colorado Health Sciences Ctr, Denver, CO
15 - Virginia Commonwealth University, Rich., VA
16 - Washington VA Medical Center, Washington
17 -Penn State Univ, Pennsylvania, PA
18 -Univ. College Dublin, Wales
19 -Yale Univ., New Haven, CT

# Appendix

1: Am J Hum Genet. 2006 Feb;78(2):315-33. Epub 2006 Jan 3.

# 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.

**Suarez BK**, **Duan J**, **Sanders AR**, **Hinrichs AL**, **Jin CH**, **Hou C**, **Buccola NG**, **Hale N**, **Weilbaecher 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**.

Department of Psychiatry and Genetics, Washington University, St. Louis, MO. pgejman@northwestern.edu.

We report the clinical characteristics of a schizophrenia sample of 409 pedigrees--263 of European ancestry (EA) and 146 of African American ancestry (AA)--together with the results of a genome scan (with a simple tandem repeat polymorphism interval of 9 cM) and follow-up fine mapping. A family was required to have a proband with schizophrenia (SZ) and one or more siblings of the proband with SZ or schizoaffective disorder. Linkage analyses included 403 independent full-sibling affected sibling pairs (ASPs) (279 EA and 124 AA) and 100 all-possible half-sibling ASPs (15 EA and 85 AA). Nonparametric multipoint linkage analysis of all families detected two regions with suggestive evidence of linkage at 8p23.3-q12 and 11p11.2-q22.3 (empirical Z likelihood-ratio score [Z(lr)] threshold >/=2.65) and, in exploratory analyses, two other regions at 4p16.1-p15.32 in AA families and at 5p14.3-q11.2 in EA families. The most significant linkage peak was in chromosome 8p; its signal was mainly driven by the EA families. Z(lr) scores >2.0 in 8p were observed from 30.7 cM to 61.7 cM (Center for Inherited Disease Research map locations). The maximum evidence in the full sample was a multipoint Z(lr) of 3.25 (equivalent Kong-Cox LOD of 2.30) near D8S1771 (at 52 cM); there appeared to be two peaks, both telomeric to neuregulin 1 (NRG1). There is a paracentric inversion common in EA individuals within this region, the effect of which on the linkage evidence remains unknown in this and in other previously analyzed samples. Fine mapping of 8p did not significantly alter the significance or length of the peak. We also performed fine mapping of 4p16.3-p15.2, 5p15.2-q13.3, 10p15.3-p14, 10q25.3-q26.3, and 11p13-q23.3. The highest increase in Z(lr) scores was observed for 5p14.1-q12.1, where the maximum Z(lr) increased from 2.77 initially to 3.80 after fine mapping in the EA families.