State of Gene Discovery Efforts

• The glass half full
  – Several replicated linkage findings (e.g., 22q11, 6p22, 8p12-21, 1q21-22, 1q42, 13q32-34, 12q24, 14q22-32)
  – Candidate genes identified in each region with varying degrees of biological plausibility (e.g., COMT, DTNBP1, NRG1, RGS4, DISC1, G72, DAAO, Akt1)
  – Some of these genes have been implicated in GWAS of rare CNVs

• The glass half empty
  – Markers and haplotypes associating with the syndrome are not consistent across studies
  – Functionally significant variants have not been identified (other than for COMT)
  – Very few (and none of these candidates) significant in GWAS of common SNPs despite large N’s
Today’s Talk

- Dissecting phenotypic & genetic issues that limit the power of *case-control* GWAS
  - What phenotype(s) do the genes encode?
  - Heterogeneity of risk alleles
- A translational strategy based on endophenotypes
  - Validation of endophenotypic traits in discordant twins
  - Association of genetic variants with endophenotypes
  - Validation of genetic associations in mutant mice
  - Examples: DISC1, Dysbindin

<table>
<thead>
<tr>
<th>Phenotype Strategy</th>
<th>Genotype Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole-Genome</td>
<td>Candidate Gene</td>
</tr>
<tr>
<td>Qualitative</td>
<td>Case-Control GWAS</td>
</tr>
<tr>
<td></td>
<td>Case-Control custom panel</td>
</tr>
<tr>
<td>Quantitative</td>
<td>QTL GWAS</td>
</tr>
<tr>
<td></td>
<td>QTL custom panel</td>
</tr>
</tbody>
</table>
Is schizophrenia truly a *categorical* phenotype?

- Multiplicity of ‘small effect’ risk factors (genetic and environmental), and absence of ‘major effect’ risk factors, predicts continuum of risk.
Liability-Threshold Model

- Liability to disorder is distributed continuously in population
- Phenotypic severity derives from liability continuum
- Threshold for disorder defined on functional/pragmatic grounds
- Sub-syndromal degrees of affection are common

\[ L_i = A_i + B_i + X_i + (A_i \times B_i) \ldots + C_i \]
What phenotype(s) do the genes that predispose to schizophrenia code for? (part 2)

- Schizophrenia does not boil down to a single symptom, cognitive or emotional state, neural system, cell signaling pathway, protein, or gene.
- Rather, schizophrenia is a syndrome involving many different symptom combinations.
- Any given case thus reflects an aggregation of disruptions in multiple neural systems, cell signaling pathways, proteins, and genes.
- Though there should be some common profiles, different cases would be expected to differ in the particular combinations of neural systems, cell signaling pathways, proteins, and genes involved.
Psychiatric Genetics: Populating the branches

Disrupted dopamine-glutamate interactions

DISC1

synaptic density

DA catabolism

Dysbindin

GRM3

COMT

Negative Symptoms & Deficits in Functioning

Working memory dysfunction

Schizophrenia

Evolutionary paradox of schizophrenia

- Schizophrenia is associated with reduced reproductive success, but is also maintained at relatively high prevalence in the population
  - E.g., Fertility Ratio (FR) = 0.39; 95% CI = 0.35-0.44 (Stahl & McCabe 2011)
- If schizophrenia decreased Darwinian fitness in ancestral environments, why hasn’t natural selection eliminated the alleles that predispose to it?
Fitness related trait

Furthest downstream / more broadly defined processes
Common vs. Rare Variants

- **Common variants**
  - For diseases that are highly heritable and relatively common, genetic susceptibility is thought to involve common variations of many genes, each of small effect
  - Typically, these are measured by SNPs, reflecting single-base pair mutations
  - Of the 0.2% genetic variation in humans, .08% is by SNPs

- **Rare variants**
  - Cytogenetic abnormalities and copy number variants (CNVs) are rarer, but potentially more highly penetrant for the trait than the common variants (SNPs)
  - These involve mutations of larger segments (typically > 100 kb)
  - Of the 0.2% genetic variation in humans, .12% is by CNVs
Evolutionary Implications

• For common variants, low selection pressure against any one gene
  – Still, any non-zero degree of selection pressure will gradually eliminate risk alleles from population
  – For polygenic traits, this process is slow, so proportion of genetic variation due to old mutations is larger than the proportion due to new mutations
  – Each person contains 500-2000 old slightly deleterious coding mutations†

• For CNVs, if they do have a moderate to large effect on illness, these are likely to be under substantial negative selection
  – These mutations are therefore rare and more frequently de novo
  – 2-3 new non-lethal but deleterious coding mutations per person*

• Against the backdrop of a relatively large pool of mutations from ancestral populations, the constant introduction of new mutations (both SNPs and CNVs) ensures heterogeneity in the particular risk alleles in genes associated with fitness-reducing phenotypes

†Fay et al, 2001; Sunayaev et al, 2001
*Keighley & Eyre-Walker, 2000
Genetic Risk for Schizophrenia:
Many Common Variants vs Few Genes of Large Effect
(Nature volume 460, August 2009)

• “Common variants conferring risk of schizophrenia” Stefansson et al., pg 744-47
  - SGENE-plus sample: 2,663 cases, 13,498 controls from 8 European locations
  - 314,868 markers

• “Common polygenic variation contributes to risk of schizophrenia and bipolar disorder” International Schizophrenia Consortium, pg 748-52
  - ICS sample: 3,322 European cases, 3,587 controls
  - 739,995 SNPs

• “Common variants on chromosome 6p22.1 are associated with schizophrenia” Shi et al., pg 753-57
  - MGS sample: 2,681 cases 2,653 controls (European) plus 1,286 cases, 973 controls (African-American); 671,424 SNPs for Europeans; 811,340 for AAs
Summary of GWAS Findings

• Even with large samples, none of these studies alone could find a genetic marker significant across the whole genome

• All 3 meta-analyses (each combining data across 3 studies in different ways) point to a region on chromosome 6 (Major Histocompatibility Complex)

• One meta analysis points to two new genes- NRGN on 11q24.2 and TCF4 at 18q21.2
Meta-analyses ISC group

• Modeled a risk “score” based on total number of alleles at different significance thresholds
  – Replicated across independent samples
  – Shared across SZ and BP
  – Specific to psychiatric phenotypes

• In simulations, models with a large number of common variants (rather than a few common variants of large effect, or rare variants) fit the observed data significantly better

Analysis stratified by score allele frequency.
Summary of design issues for case-control GWAS

- Case-Control GWA studies model as a categorical phenotype a set of intrinsically quantitative phenomena

- Syndromal phenotype is the furthest downstream outcome, thus subject to the greatest heterogeneity
  - On one hand, case-control GWAS should thus be best positioned to detect old mutations (SNPs) that have spread the furthest in human populations
    - Supported by replication of genome-wide risk “score” across independent samples in ISC
  - On the other hand, those mutations probably have very small effect sizes on the probability of illness (or would have already been eliminated)
    - Supported by paucity of alleles reaching $10^{-8}$ significance in samples of +10,000
A Strategy for Gene Discovery & Confirmation Based on Endophenotypes

- Identify **intermediate phenotypes** (e.g., using brain imaging, neurocognitive testing, gene expression profiling) that co-vary in a dose-dependent manner with genetic loading in twin pairs *discordant for illness phenotype* (MZ co-twins > DZ co-twins > normal twins)
  - Use of unaffected twins from discordant pairs disconfounds genetic from phenotypic effects
- Search for genetic polymorphisms (both common and rare) that contribute to quantitative variation in these endophenotypes
- Evaluate transgenic models on panel of homologous endophenotypes in mice
Plasticity-Related Endophenotypes

Short- and Long-Term Memory Function*

Gray Matter in PFC & MTL*

Reduced spine density

* Validated using MZ and DZ twins discordant for SCZ and healthy controls (disconfounds genetic influences from secondary factors)


Glanz & Lewis *Arch Gen Psy* 2000;57:65-73.

• DISC1 discovered via a balanced translocation

• DISC1 expressed in brain - especially in hippocampus and cortex

• In vitro truncation of DISC1 (inhibiting its interaction with NUDEL) inhibits neurite outgrowth

• DISC1 thus a candidate for explaining disruptions in synaptic plasticity and connectivity in schizophrenia
**DISC1 Translational Analyses**

<table>
<thead>
<tr>
<th>DISC1 Haplotype</th>
<th>Asociality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>DISC1 HEP2/3</td>
<td>96%</td>
</tr>
<tr>
<td>AATG</td>
<td>83%</td>
</tr>
</tbody>
</table>

**Human** (Cannon et al. Arch Gen Psychiatry 2005) | **Mouse** (Li et al. PNAS 2007)

- DISC1 HEP2/3 associates with social deficits
- DISC1 HEP2/3 and HEP1 associate with deficits in spatial working memory and long-term memory
- DISC1 HEP2/3 and HEP1 associate with reductions in gray matter in prefrontal cortex and hippocampus
- Disc1-cc results in reduced sociality
- Disc1-cc results in reduced spatial working memory
- Disc1-cc results in reduced dendritic complexity and basal synaptic transmission of hippocampal neurons
Dysbindin

• Best candidate under a replicated linkage peak at 6p22.3 (Straub 2002)

• A number of dysbindin SNPs and haplotypes associate w/ schizophrenia (though inconsistently across study samples)

• Functional variants remain unknown, though risk may be associated with low expression of mRNA and protein (Bray et al. 2005; Tang et al. 2009)

• Robust associations of dysbindin SNPs with impaired cognition

• Gene encodes dystrobrevin-binding protein: classically thought to be a member of BLOC-1, also known to be involved in glutamate vesicle trafficking to synapse
Reduced NMDAR-mediated synaptic plasticity in schizophrenia

- PCP & ketamine, which block NMDA receptors, induce/exacerbate symptoms of schizophrenia (+, -, cognitive)

- Several studies have observed decreased NMDAR function in schizophrenia, including in drug-free patients

Pilowsky et al., *Mol Psychiatry* 2006;11:118-119
Dysbindin Translational Analyses

- Null mutant model arose spontaneously on DBA/2J background (sdy mouse); backcrossed to C57B1
- Mutants show disrupted spatial WM, reduced activity in glutamatergic neurons, reduced expression of NR1 subunit of NMDA receptor, and reduced LTP (rescued with bath application of glycine)
Manganese-Enhanced Contrast
Wildtype > Dysbindin -/-

havenuela, dentate, CA1, CA2/3

Tensor-Based Morphometry
Dysbindin -/- > Wildtype

bilateral auditory cortex, thalamus, and dentate gyrus

Lutkenhoff et al. Under review
Are DISC1 and Dysbindin susceptibility genes for schizophrenia?

• Linkage
  – Pretty solid statistical evidence of linkage to both regions, replicated across samples and gene pools
  – However, linked regions are relatively large and thus neighboring genes could drive the linkage findings

• Association
  – No SNPs in either gene have reached genome-wide significance in case-control GWA studies
  – Rare CNVs in each gene have been seen in a few cases
  – High likelihood of multiple mutations in both genes in our ancestral past
  – No reason to assume same mutation is relevant to every case

• Biological plausibility
  – SNP associations to validated endophenotypes for schizophrenia
  – Mutant mice show deficits on analogous endophenotypes
Comprehensive Candidate Gene Study

- Examined 3,126 markers from 50 genomic regions based on meta-analysis of all published schizophrenia genetic association studies (www.szgene.org)
  - HuGENet (Human Genome Epidemiology Network) interim criteria based on sample size, heterogeneity across studies, and protection from bias
- Initial series of 135 subjects were scanned & genotyped
  - 26 MZ twin pairs, 34 DZ twin pairs, 15 individuals
  - Status: 96 healthy controls, 26 SZ, 13 BP
  - Mean age = 52 years (SD=10)
- Screened markers for association with global brain volume, then tested regional effects with Tensor Based Morphometry (TBM)
## Genomic Regions

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Genes (# SNPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHR 1</td>
<td>TSNAX (10), DISC1 (181), PDE4B (196), MTHFR (47), RGS4 (9), GRIK3 (78), PLXNA2 (127), GSTM1 (4), IL10 (9)</td>
</tr>
<tr>
<td>CHR 2</td>
<td>GAD1 (14), ZNF804A (65), IL1B (9)</td>
</tr>
<tr>
<td>CHR 4</td>
<td>CCKAR (6)</td>
</tr>
<tr>
<td>CHR 5</td>
<td>GABRB2 (63), CMYA5 (93)</td>
</tr>
<tr>
<td>CHR 6</td>
<td>TNF (15), PRSS16 (44), <strong>PGBD1</strong> (41), NOTCH4 (106), HIST1H2BJ (8), AHI1 (47), RPP21 (22), DTNBPI (40), C6orf217 (47), MDGA1 (57)</td>
</tr>
<tr>
<td>CHR 7</td>
<td><strong>RELN</strong> (271)</td>
</tr>
<tr>
<td>CHR 8</td>
<td>NRG1 (388), PPP3CC (28), SLC18A1 (28)</td>
</tr>
<tr>
<td>CHR 10</td>
<td>GWA_10q26.13 (1)</td>
</tr>
<tr>
<td>CHR 11</td>
<td>NRGN (4), DRD4 (3), DRD2 (29), GWA_11p14.1 (1), TPH1 (13), OPCML (251)</td>
</tr>
<tr>
<td>CHR 12</td>
<td>GRIN2B (205), DAO (5)</td>
</tr>
<tr>
<td>CHR 13</td>
<td>DAOA (18), HTR2A (57)</td>
</tr>
<tr>
<td>CHR 14</td>
<td>AKT1 (14)</td>
</tr>
<tr>
<td>CHR 16</td>
<td>RPR1P1L (29), HP (1)</td>
</tr>
<tr>
<td>CHR 17</td>
<td>SRR (7)</td>
</tr>
<tr>
<td>CHR 18</td>
<td>IMPA2 (36), TCF4 (83)</td>
</tr>
<tr>
<td>CHR 19</td>
<td>APOE (9)</td>
</tr>
<tr>
<td>CHR 22</td>
<td>PRODH (20), ADRBK2 (32), COMT (49)</td>
</tr>
</tbody>
</table>
Genotyping Quality Control

- 19 SNPs excluded for failing Hardy-Weinberg Equilibrium (HWE) test (p<0.001)
- 568 SNPs excluded for minor allele frequencies (MAF) < 0.01
- 7 SNPs excluded for frequency of missing data > 0.05
- No SNPs excluded for zygosity checks, genotype frequency and category checks

- 6 individuals excluded by population stratification test (EIGENSTRAT)
- 2 individual excluded by inbreeding test (too much homozygosity)

Final Data:
- 2345 total markers in 50 genes
- 127 individuals
• **PLINK: QFAM** (Shaun Purcell et al, 2007 AJHG)
  - Family-based association tests for quantitative traits
  - Based on QTDT package (Fulker et al, 1999, AJHG and Abecasis et al, 2000, AJHG)
  - Accounts for twinship, but not completely for MZ twins:
    - 1 co-twin excluded from each MZ twin pair for analysis
    - Then analysis re-run with other co-twin from each MZ twin pair excluded
  - 2 step permutation:
    1. QFAM total (between and within components) with adaptive permutation on all quality controlled marker
    2. QFAM total with max(T) 100,000,000 permutations on significant markers from 1st step
Multiple Testing Correction

• Used an adjusted Bonferroni correction that models the underlying linkage disequilibrium (LD) structure among the markers to determine the number of effectively independent tests (Nicodemus, Liu et al. 2005)
  – 2,345 total markers used in analysis
  – 941 SNPs in independent LD blocks

• Corrected p value for 941 tests:
  – \( p = 0.05 / 941 = 5.31 \times 10^{-5} \)

• Therefore, p-values less than 5.31 \( \times 10^{-5} \) are statistically significant (p-values less then 10\(^{-4}\) considered suggestive)

• P-values were empirically derived using permutation
RELN: rs262366

RELN: Reelin
Chromosome 7
rs262366 located in intron

<table>
<thead>
<tr>
<th>rs262366</th>
<th>Brain</th>
<th>ICV</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZ 1</td>
<td>9.25E-05</td>
<td>4.19E-05</td>
</tr>
<tr>
<td>MZ 2</td>
<td>1.17E-04</td>
<td>2.70E-05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rs262366</th>
<th>BP</th>
<th>SZ</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>2</td>
<td>10</td>
<td>32</td>
</tr>
<tr>
<td>GA</td>
<td>20</td>
<td>11</td>
<td>60</td>
</tr>
<tr>
<td>GG</td>
<td>10</td>
<td>6</td>
<td>22</td>
</tr>
</tbody>
</table>

Major allele = A
Minor allele = G
MAF = 0.4945
Orthographic views of EMMAX modeling for effect of RELN polymorphism rs262366 on TBM maps; significantly associating voxels in red and yellow located in the superior frontal gyrus and anterior cingulate gyrus (515 voxels, FDR corrected; p < 1.24 x 10^{-5}). Three-dimensional volume rendering of representative brain with rs262366 association result displayed in blue (lower right).
rs262366 (RELN): No threshold
Reelin & Schizophrenia

SZGene meta-analysis for RELN (rs7341475): A vs. G

<table>
<thead>
<tr>
<th>Study</th>
<th>OR</th>
<th>95% CI</th>
<th>I²</th>
</tr>
</thead>
<tbody>
<tr>
<td>All studies 0</td>
<td>0.89</td>
<td>[0.83,0.97]</td>
<td>0</td>
</tr>
<tr>
<td>Caucasian studies</td>
<td>0.88</td>
<td>[0.81,0.96]</td>
<td>0</td>
</tr>
<tr>
<td>Study specific ORs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ben-David, 2010 [O] †</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Tost, 2010 [O]</td>
<td>1.64</td>
<td>[0.72,3.72]</td>
<td>–</td>
</tr>
<tr>
<td>Liu, 2010 [C]</td>
<td>0.93</td>
<td>[0.80,1.08]</td>
<td>–</td>
</tr>
<tr>
<td>Shifman, 2008, China [A] •</td>
<td>1.03</td>
<td>[0.74,1.42]</td>
<td>–</td>
</tr>
<tr>
<td>Shifman, 2008, Ireland [C] •</td>
<td>0.90</td>
<td>[0.73,1.11]</td>
<td>–</td>
</tr>
<tr>
<td>Shifman, 2008, Israel [C] •</td>
<td>0.85</td>
<td>[0.73,0.98]</td>
<td>–</td>
</tr>
<tr>
<td>Shifman, 2008, UK [C] •</td>
<td>0.87</td>
<td>[0.71,1.06]</td>
<td>–</td>
</tr>
<tr>
<td>Shifman, 2008, USA (NIMH) [C] •</td>
<td>0.86</td>
<td>[0.67,1.11]</td>
<td>–</td>
</tr>
</tbody>
</table>

All control populations in HWE (P>0.05)

• Initial study  † No data provided, or data not eligible for inclusion  0 Meta-analysis after excluding initial study n.a.

A Asian  C Caucasian  O Other/Mixed

Chr 7

[Genome-wide association study results visualized on a chromosome map, showing regions associated with REELN (rs7341475).]
Reelin Linkage Disequilibrium
Reelin & Neural Development

Wildtype

Reelin Mutation

Cece2309. Wikimedia Commons. 2007.
Conclusions

• Imaging phenotypes can be informative for genetic studies of complex illness
  – Quantitative, multidimensional, and translational nature represent significant advantages

• In schizophrenia, some utility of this approach in relation to some candidate genes, including DISC1, dysbindin, reelin

• Power and sample size the biggest rate limiting step both at the genetic (SNP) and phenotypic (voxel) levels for both candidate gene and GWA
Contributors

- Frank Sun
- Theo van Erp
- Katherine Karlsgodt
- Evan Lutkenhoff
- Matthew Keller
- David Glahn
- Arthur Toga
- Paul Thompson
- Jason Stein
- Alcino Silva
- Weidong Li
- David Jentsch
- Antoinetta Lavin
- David Glen
- Matti Huttunen
- Jouko Lonnqvist
- Jaakko Kaprio
- Leena Peltonen
- Jesper Ekedland
- William Hennah
- Tiina Paunio

Grant Support: National Institute of Mental Health (MH52857, MH081902, MH083269), National Center for Research Resources, Staglin Family, Music Festival for Mental Health