Methods in Target Identification: A System View

Suzanne Brewerton
Overview

• Why take a systems view?
• Psychiatric Genetics Consortium
• Data from Genome Wide Association Studies
• Ranking Targets from Genetics
• Network analysis of GPCRs in Schizophrenia
• Systems biology collaboration
Identification of pathway and target involved in disease

Assumption:
One target → One consequence

In reality:
The target is usually one component of a complicated biological network

- A target acting as a single critical node may control or influence many processes
- Network interactions can show redundancy. “Work arounds” limit efficacy of a drug
- Drugs usually interact with multiple targets
- Efficacy and safety are often a consequence of interaction with multiple targets
Example Alzheimer’s (LOAD) Network


“This is too big; I don’t know what to do with it.”
Reduce complexity

Figure key:

Five main immunologic families found in Alzheimer’s-associated module

Square nodes in surrounding network denote literature-supported nodes.

Core members of different immunologic families are shaded darkly.

(Interior circle) Width of connections between 5 immune families are linearly scaled to the number of inter-family connections.

Labeled nodes are either highly connected in the original network, implicated by at least 2 papers as associated with Alzheimer’s disease, or core members of one of the 5 immune families.

Node size is proportional to connectivity in the full module.

Drugs often exhibit polypharmacology, binding multiple targets.

The chart shows the known affinity (Ki) values of antipsychotic drugs for a panel of receptors.

Accelerated Genetic Discovery in Psychiatry

Rare CNV GWAS

0 50 100 150 200 250

Genetics of Complex Diseases

- A large and growing number of Mendelian and chromosomal disorders have been assigned to particular causal genetic events.
- Complex diseases are caused by a combination of genetic, environmental, and lifestyle factors, most of which have not been identified.
- Complex diseases do not obey the standard Mendelian patterns of inheritance. They do not always develop despite the presence of susceptibility gene(s), i.e. they show incomplete penetrance and variable expressivity.

<table>
<thead>
<tr>
<th>Birth Defects: cleft palate/lip, neural tube defects such as spina bifida</th>
<th>Cancer: bowel, breast, ovarian, bowel, melanoma and prostate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular conditions: high blood pressure, some causes of heart disease, high cholesterol</td>
<td>Metabolic: diabetes</td>
</tr>
<tr>
<td>Neurological/psychiatric conditions: Alzheimer disease in later life, schizophrenia, bipolar disorder</td>
<td>Muscular/skeletal: arthritis, rheumatic disorders, osteoporosis</td>
</tr>
<tr>
<td>Skin conditions: psoriasis, moles, eczema</td>
<td>Respiratory: asthma, allergies, emphysema</td>
</tr>
</tbody>
</table>
“There are very few new molecular entities, very few novel ideas, and almost nothing that gives any hope for a transformation in the treatment of mental illness.”

— Thomas Insel, M.D., Director NIMH

Psychiatric Genomics Consortium (PGC)

What Is The PGC?

- Conduct *meta-analyses of genome-wide genetic data for psychiatric disease.*
- Confederation of most investigators in the field.
  - PGC scientists are from over 60 institutions in 19 countries.
  - This is the largest consortium in the history of psychiatry, and the largest biological experiment in the field.
- Initial intent to investigate the *common single nucleotide polymorphisms (SNPs)* genotyped on commercial arrays. Expanded to include structural variation (copy number variation) and uncommon or rare genetic variation.
- The basic idea:
  - Individual studies are generally too small to identify robust and replicable associations
  - Meta-analysis is a widely-used technique that can combine information across studies.
- Focused on five critically-important disorders: *autism, attention-deficit hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia.*
  - Added anorexia nervosa, OCD/Tourette's, and PTSD in 2013.
Genome Wide Association Studies (GWAS)

- Measures and analyzes DNA sequence variations from across the human genome.
- Identify genetic risk factors for diseases that are common in the population.
- Chip-based microarray technology for assaying one million or more SNPs.
- Typically, there are two commonly occurring base-pair possibilities for a SNP location.
- Allele frequency of a SNP is the frequency of the less common allele (minor allele).
- The odds ratio is used to quantify how strongly the presence or absence of property A is associated with the presence or absence of property B in a given population.
Genomic coordinates are displayed along the X-axis, with the negative logarithm of the association $P$-value for each single nucleotide polymorphism displayed on the Y-axis.

It gains its name from the similarity of such a plot to the Manhattan skyline: a profile of skyscrapers towering above the lower level "buildings" which vary around a lower height.
Schizophrenia GWAS

SCZ – Ancient History
2601 cases, 3345 controls
0 genome wide significant sites

PGC SCZ – The Past – 2011
9394 cases, 12462 controls
5 genome wide significant sites

PGC SCZ – 2012
25785 cases, 28441 controls
62 genome wide significant sites

PGC SCZ – June 2013
35476 cases, 46839 controls
97 genome wide significant sites
• PGC2 main GWAS data analysis complete - 36,989 schizophrenia cases (submitted March 2014)

• **128 independent associations** spanning 108 conservatively defined physically distinct loci that meet genome-wide significance

• “COMMON VARIANT ASSOCIATION META-ANALYSIS FOR SCHIZOPHRENIA IDENTIFIES 108 GENOMIC LOCI AND IMPLICATES POSTSYNAPTIC AND IMMUNE PROCESSES”

• Associations to DRD2 (D2 receptor), multiple voltage-gated calcium channels, and genes involved in glutamatergic neurotransmission.

• Genes with … *expression in brain*, and particularly *neurons* rather than glia, were enriched for associations, pointing to an important neuronal pathology.

• Pathway analyses identified enrichment of associations to genes involved in synaptic transmission, principally *dendritic spines* and the *post-synaptic density*.

• Associations were enriched among *genes expressed in tissues that play important roles in immunity*
Genetic Target Ranking

Genetic association ranking

- GWAS ranked score
- Functionality of variants
- SNP to target mapping

Tractability

- Ligandability/druggability
- Biopharmability
- Internal tools and assets
- External resources

Rare variants and cross disorder association

- Exomes and genomes
- Copy number variants
- Rare disorder genes
- Cross disorder GWAS

Toxicology

- Known on target toxicity
- Close target toxicity
- Tissue distribution
- Mouse phenotypes
- Other disease assoc.

Preclinical qualification

- Brain expression patterns
- Expression changes in disease
- Mouse phenotypes
- Biology and pharmacology
- Text mining

Pathways and Networks

- Biochemical pathways
- Genetic pathway analysis
- Gene expression networks
- Systems biology networks
Pathway and Network Analysis

Genetic association data

Gene expression modules in schizophrenia brain

Protein interaction networks

Integration

Disease pathology networks
Network Analysis

GPCR Subfamily Members → Thompson Reuters MetaCore Interaction Data → Schizophrenia GWAS gene list

GPCR to Schizophrenia Subfamily networks → Merged GPCR to Schizophrenia network

Network hubs

Network statistics calculated using CytoScape, CentiScape

Thompson Reuters MetaCore Interaction Data

- Literature derived interaction data
- Manually assembled and curated
- Pathway type data, includes binding, cleavage, covalent modification, phosphorylation, transport, transformation, catalysis, transcription regulation etc. ...

Combined ChEMBL and Lilly dataset, activity cutoff <1uM
Dopamine signalling in Schizophrenia

2 step shortest paths network

From: Dopamine receptor subtypes

To: Proteins defined by PGC2 schizophrenia GWAS genes

- Schizophrenia GWAS genes
- Dopamine receptor subtypes
GPCRs interact with schizophrenia GWAS genes

- Adenosine receptor network
- Dopamine receptor network
- 5HT receptor network
- Histamine receptor network
- Adrenergic receptor network
- Opioid receptor network
- Metabotropic glutamate receptor network
GPCRs interact with schizophrenia proteins

- Merged 7 networks describing 2 step shortest paths from GPCR subfamilies to proteins derived from genes significantly associated with schizophrenia in PGC2 GWAS data

Nodes – proteins
Edges – biological interactions

Nodes sized by degree (number of interactions)
Evidence for hub gene criticality in molecular networks

Lethality is strongly associated with centrality in S. cerevisiae, E.coli, D.melanogaster and C.elegans

Nature 411, 41-42 (3 May 2001)
Hubs in GPCR schizophrenia network

- Proteins identified by schizophrenia GWAS
- GPCRs and GPCR signaling proteins

Remove scz GWAS data
Remove proteins degree <10

Hubs involved in GPCR signalling in schizophrenia
Compounds target several network hubs

- *Hubs in GPCR schizophrenia network*
- *Integrated Lilly and ChEMBL data*
- *Activity cutoff 1uM*

Nodes – proteins
Edges – compounds in common

Nodes sized by increasing degree (number of interactions)
Edges sized by increasing number of compounds
Systems biology through collaboration

**IGMB**
Generation of integrative network models of schizophrenia

**Lilly**
design cassettes of compounds to interrogate disease sub-networks in an informed and hypothesis-driven way

**Lilly**  **IGMB**
- Discovery of novel target strategies for schizophrenia, especially MTDD
- Understanding the mechanism of action (MoA) of drugs
- New molecular markers of therapeutic response
- Information on potential toxic effects

**IGMB**  **Lilly**
Model and characterize schizophrenia network drug response using drug transcriptional profiles
Summary

• Putting an emphasis on genetic data in target identification
• Expanding capabilities through collaboration
• Using systems approaches to understand
  • The key regulators in schizophrenia
  • How to use polypharmacology of compounds
Acknowledgements

• David Collier (Psych Genetics)
• Brian Eastwood (Statistics)
• Younes Mokrab (Psych Genetics)
• James Scherschel (Research IT)
• Cara Ruble (TTx)
• Peng Yu (Research IT)
• Mike Bodkin (CADD)
• David Evans (CADD)
• Derek Marren (Research IT)
Association between ESR1 gene polymorphisms and haplotypes with schizophrenia.
Wang LW, He SP, Zhang JL, Tong DX, He Y, Cheng XL.
Chinese.

Gender-specific reduction of estrogen-sensitive small RNA miR-30b in subjects with schizophrenia.
Mellios N, Galdzicka M, Ginns E, Baker SP, Rogaev E, Xu J, Akbarian S.

A functional polymorphism in estrogen receptor alpha gene is associated with Japanese methamphetamine induced psychosis.

Variants in the estrogen receptor alpha gene and its mRNA contribute to risk for schizophrenia.

Analyses of variants located in estrogen metabolism genes (ESR1, ESR2, COMT and APOE) and schizophrenia.
Disrupted-in-**Schizophrenia**-1 expression is regulated by beta-site amyloid precursor protein cleaving enzyme-1-neuregulin cascade.

**The beta-secretase enzyme BACE in health and Alzheimer's disease: regulation, cell biology, function, and therapeutic potential.**
Vassar R, Kovacs DM, Yan R, Wong PC.

**Levels of [(3)H]pirenzepine binding in Brodmann's area 6 from subjects with schizophrenia is not associated with changes in the transcription factor SP1 or BACE1.**
Dean B, Soulby A, Evin GM, Scarr E.
Schizophrenia susceptibility pathway neuregulin 1-ErbB4 suppresses Src upregulation of NMDA receptors.
Pitcher GM, Kalia LV, Ng D, Goodfellow NM, Yee KT, Lambe EK, Salter MW.

FYN kinase gene: another glutamatergic gene associated with bipolar disorder?

[Analysis of the fyn kinase gene in Alzheimer's disease and schizophrenia].
Approach

1. Expand 2107 Index SNPs to full list ✓

2. Map each variant to a(several) gene(s)
   - in coding region of a gene ✓
   - non-coding region of a gene ✓
   - proximal to a gene (“nearest gene”) ✓

3. Determine number of independent and total hits ✓

4. Score based on number and type of independent hits ✓

5. Construct appropriate plots to accompany results
Basic Dataset Facts

Numbers
- 145095 different variant-gene observations
- 112240 unique variants-entries
- 24275 have multiple entries due to multiple gene mappings per variant
- all results are sig at 10-4 or lower.
- stated OR are for named variant. All are restated to the less-frequent variant in the control sample

Imputation Quality
- 144113 are categorized as HQ2
- 982 are blank (not HQ2)

Gene Mapping
- 42355 do not map to a known Gene symbol or Ensemble Gene number
-102,739 variants map to named Genes

Genes
• 5735 unique Ensemble Genes
• 5604 unique named genes
Outcomes

Gene Site Categories

Coding
• exon
• exon_and_intron

Non-Coding
• intron
• utr

Proximal
• outside_gene_tbs
• outside_gene_regulatory
• outside_gene_5KB
• outside_gene_intergenic

Statistical Sig Categories
• LowSig $1 \times 10^{-6} < p < 1 \times 10^{-4}$
• ExplSig $5 \times 10^{-8} < p < 1 \times 10^{-6}$
• GWASsig $p < 5 \times 10^{-8}$

For GWAS and Exploratory sig only
• number of proximal hits
• number of non-coding hits
• number of coding hits
• average AbsOR* of non-coding hits
• max AbsOR of coding hits

GWAS-Sig coding hits only
• directions: $\text{OR}<1$ and $\text{OR}>1 = 2$
  others = 1

*AbsOR = max(OR, 1/OR)

Then summarize by Gene
Scoring Algorithm Version 3

Coding Score:
- 1000 points for each GWAS sig hit. BScore: 50% bonus if bidirectional
- 333 points for each Expl-sig hit. BScore: 50% bonus if bidirectional
- 100 points for each Low-sig Hit.

Non-coding Score:
- 100 points for each GWAS sig hit.
- 33 points for each Expl-sig hit.
- 10 points for each Low-sig Hit.

Proximal Score:
- 10 points for each GWAS sig hit.
- 3.3 points for each Expl-sig hit.
- 1 points for each Low-sig Hit.

Total [Dep Ignored] Score = Coding Score + Non-Coding Score + Proximal Score
Total [Dep Ignored] BScore adds 50% bonus where indicated above
Network Analysis

- A platform approach – GPCRs
- Using Thompson Reuters Metacore identify the shortest paths interactions ‘from’ GPCR subfamily members ‘to’ schizophrenia GWAS genes
  - Genes defined by PGC2 GWAS associated with SNPs (p-value ≤ 10^{-4})
- Calculate network statistics (Cytoscape, CentiScaPe)
  - Degree, Radiality, Closeness, Stress, Betweenness
- Identify hubs in the network
- Map Lilly and ChEMBL Compounds to the network hubs
Basics of building gene coexpression networks

Linear relationships between expression fluctuations defined as coexpression links

Genome-wide networks created by correlating all combinations of genes

Sage Bionetworks
1. Building normal networks

![Diagram of gene interactions]

---

Lee et al. Sage Bionetworks
2. Prioritise the disease relevance of modules

RNAseq/Array analysis in human brain over development and in disease

Genome-wide expression networks in schizophrenia versus control brain

Roussos et al.
Infer causal relationships

BrainCloud allows researchers to extract focused individual gene data from within vast genome-wide gene expression and SNP polymorphism datasets.

**Sage Bionetworks**

**Goal:** Use eSNPs as causal anchors to infer causal expression relationship between genes

**A. True directed relationships between genes**

**B. Observed correlations (bi-directional & indistinguishable)**

**C. Incorporate eSNPs as causal anchors**

Apply BIC causality test to fit network to eSNPs

**D. Fit many networks to data, select average network as most likely**

Q = genomic information (eSNP, CNV etc)

(Best fit)

(correctly inferred structure)
Hub nodes in networks
The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease

Justin Lamb,1,4 Emily D. Crawford,1,4 David Peck,3 Joshua W. Modell,1 Irene C. Blat,1 Matthew J. Wrobel,1 Jim Lerner,1 Jean-Philippe Brunet,1 Aravind Subramaniam,1 Kenneth N. Ross,3 Michael Reich,3 Haley Hieronymus,3,2 Guso Wei,5 Scott A. Armstrong,2,3 Stephen J. Haggarty,3,4 Paul A. Clennin,3 Ru Wu,3 Steven A. Carr,3 Eric S. Lander,3,5,6 Todd R. Golub3,2,5,7,7

To pursue a systematic approach to the discovery of functional connections among diseases, genetic perturbation, and drug action, we have created the first installment of a reference collection of gene-expression profiles from cultured human cells treated with bioactive small molecules, together with pattern-matching software to mine these data. We demonstrate that this “Connectivity Map” resource can be used to find connections among small molecules sharing a mechanism of action, chemicals and physiological processes, and diseases and drugs. These results indicate the feasibility of the approach and suggest the value of a large-scale community Connectivity Map project.

Creating a First-Generation Connectivity Map Perturbagen. We studied 164 distinct small-molecule perturbagens, selected to represent a broad range of activities, and including U.S. Food and Drug Administration (FDA)-approved drugs and non-drug bioactive “tool” compounds. We included multiple compounds sharing molecular targets (e.g., bispec transacylase inhibitors) to determine whether such compounds would share a molecular signature. Similarly, we profiled

is truly generalizable, systematic, and biologically-relevant. However, several potential pitfalls must be considered. Concretely, a large number of parameters would need to be optimized for each perturbation, including cell type, concentration, and treatment duration. Equally, analytical methods capable of detecting relevant signals in the data might not be generally applicable. If so, generation of a useful Connectivity Map would be impractical. However, here we demonstrate—that the Connectivity Map concept is indeed viable.

Company Confidential
Copyright © 2014 Eli Lilly and Company
Functional chemigenomics screen: Chemical perturbagens against disease networks *in silico*
Node Centralities: Degree, Betweenness, Stress

Degree
- Number of nodes directly connected to a given node
- Proteins with a very high degree are likely to be regulatory hubs

Betweenness
- Probability of a node to occur on a randomly chosen shortest path between 2 randomly chosen nodes
- The capability of a protein to bring into communication distant proteins

Stress
- The number of shortest paths passing through $n$
- The capability of a protein to hold together communicating nodes