PGC Worldwide Lab Call Details

DATE: Friday, March 13th, 2015

PRESENTER: Gerome Breen, PhD, Senior Lecturer, Lead Translational Genomics Group, MRC SGDP Centre; Genomics and Biomarkers BioResource (Biobanking) Lead; SLaM NIHR BRC for Mental Health, Institute of Psychiatry, Psychology and Neuroscience, King's College London

TITLE: "The PGC Pathway Analysis Group: Current Findings and Future Directions."

START: We will begin promptly on the hour.

1000 EDT - US East Coast
0700 PDT - US West Coast
1400 GMT - UK
1500 CET - Central Europe
0100 Sat March 14th AEDT – Australia

DURATION: 1 hour

TELEPHONE:

- US Toll free: 1 877 703.6109
- International direct: +1 617 399.5126
- Toll-free number? See http://www.btconferencing.com/globalaccess/?bid=288_attended
- Operators will be on standby to assist with technical issues. "*0" will get you assistance.
- This conference line can handle up to 300 participants.

PASSCODE: 188 641 29

Lines are Muted **NOW**

Lines have been automatically muted by operators as it is possible for just one person to ruin the call for everyone due to background noise, electronic feedback, crying children, wind, typing, etc.

Operators announce callers one at a time during question and answer sessions.

Dial *1 if you would like to ask a question of the presenter. Presenter will respond to calls as time allows.

Dial *0 if you need operator assistance at any time during the duration of the call.

UPCOMING PGC Worldwide Lab

DATE: **Friday, April 10th, 2015**
PRESENTER: Kevin Eggan, PhD, Professor in the Department of Stem Cell and Regenerative Biology, Harvard University.
TITLE: To Be Announced
START: We will begin promptly on the hour.
1000 EDT - US East Coast
0700 PST - US West Coast
1500 BST - UK
1600 CEST - Central Europe
Midnight AEST – Australia (Fri., April 10th to 0100 Sat., April 11th, 2015)
DURATION: 1 hour

TELEPHONE:
- US Toll free: 1 877 703.6109
- International direct: +1 617 399.5126
- Toll-free number? See http://www.btconferencing.com/globalaccess/?bid=288_attended
- Operators will be on standby to assist with technical issues. "*0" will get you assistance.
- This conference line can handle up to 300 participants.

PASSCODE: 188 641 29

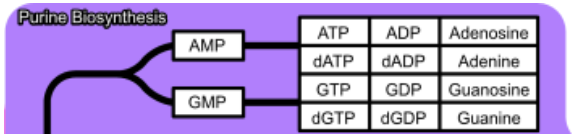
The PGC Pathway Analysis Group: Current Findings and Future Directions.

Gerome Breen PhD.

NIHR BRC for Mental Health and MRC SGDP Centre
Institute of Psychiatry
King's College London

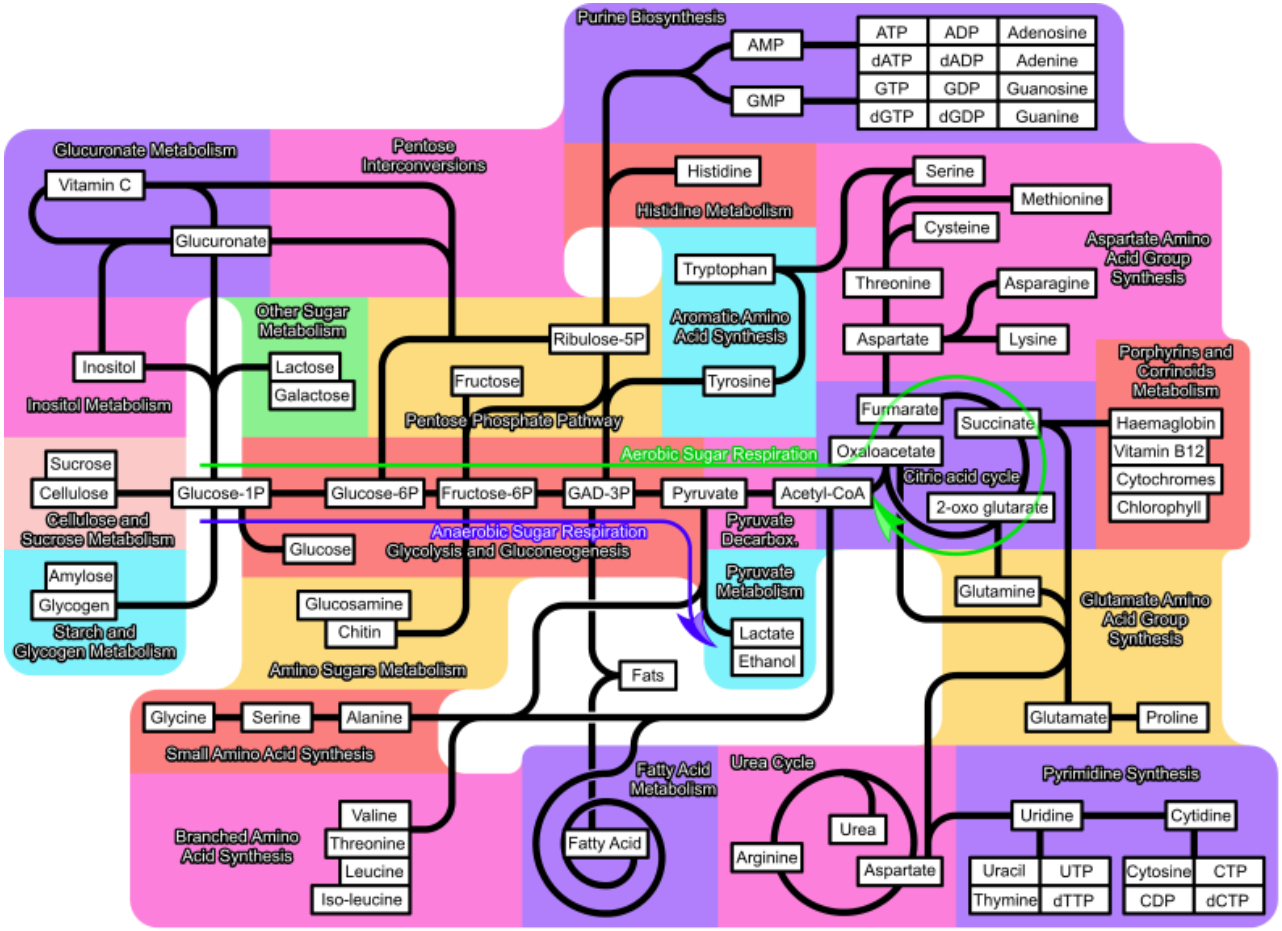
on behalf of the Psychiatric Genomics Consortium NPA group.

What are pathways?



What are pathways?

- You might say pathways don't exist….
- A view of the myriad of interactions underlying a biological process



Metabolism interaction pathway courtesy of Wikipedia

Tension of Opposites: Manually Curated vs HT Experimental

Human Protein Interaction Map

The nodes are all known protein

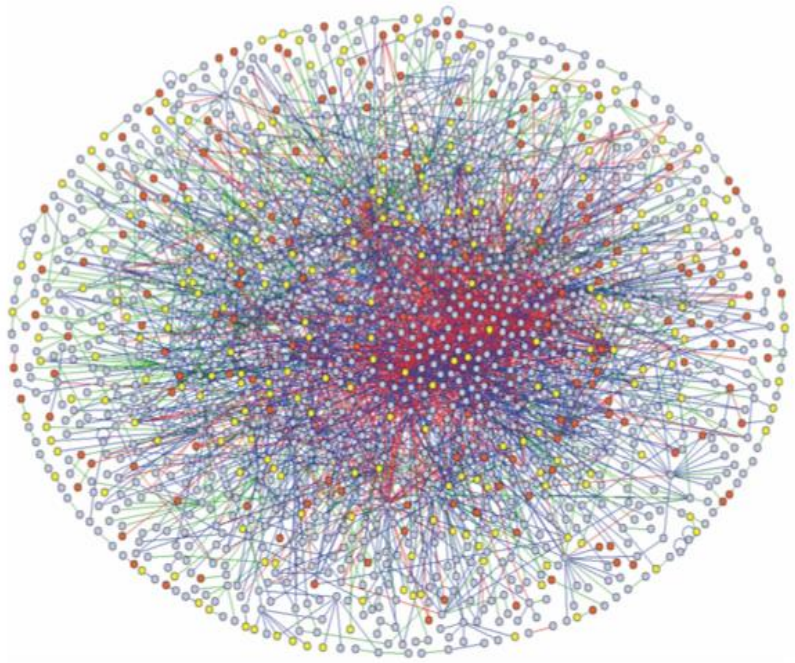
The links stand for physical binding of pairs of proteins

Screened by yeast 2 hybrid interaction mating

A group of proteins may form a functional protein complex

1705 Proteins

3186 Interactions



Stelzl *et. al*. **A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome.** Cell 2005 122: 957-968

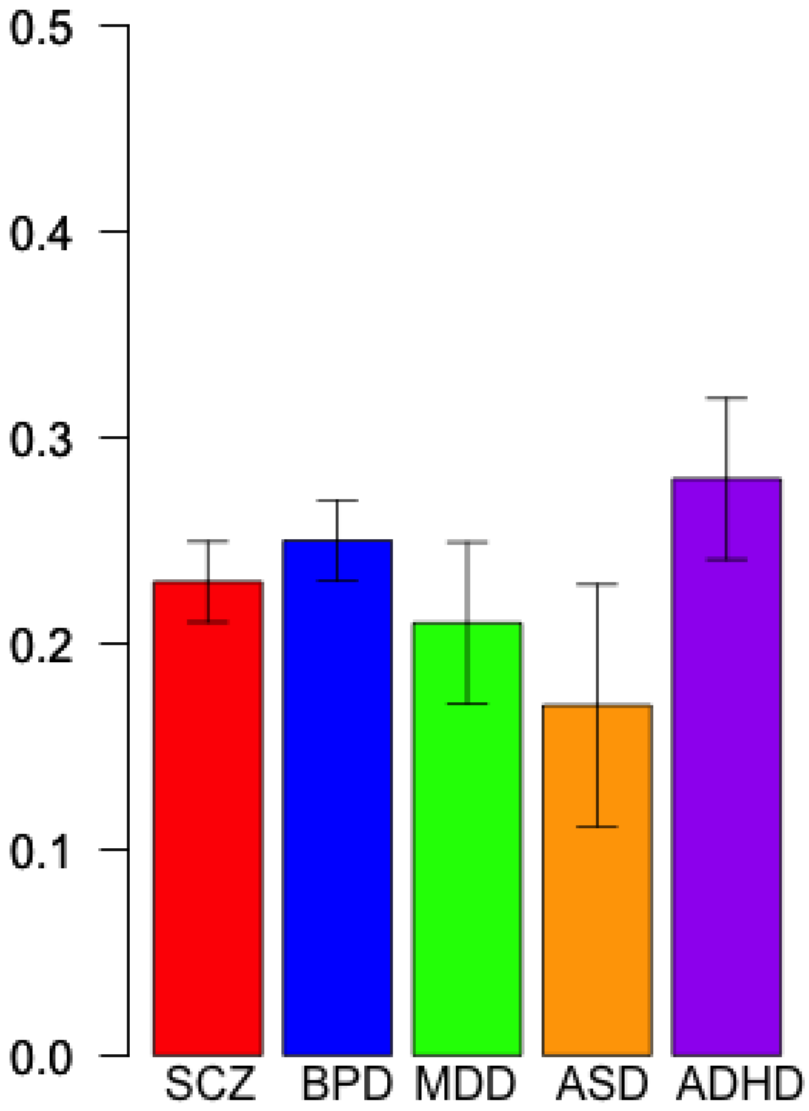
Pathway resources

- Public tools
- Curated
- Automated
- Linking up multiple resources
- Text mining
- Available as part of a company website

•Sigma Aldrich PathFinder

Image from GeneGO.com

SNP-chip-heritability



<<half twin heritability captured by GWAS (except perhaps for MDD)

Epistasis – statistically impossible to detect with current sample sizes.

Epigenetics – perhaps

Phenotype definition – yes, in psych disorders early onset is >>environmental than older cases.

Untagged rare variants – yes but is the increase in effect size enough to overcome loss of power.

De-novo variants – yes. Many examples in Autism, Schizophrenia but small % variance explained.

Biology – forget all that – what is the underlying biology of the thousands of variants causing this amount of heritability.

A Note of Caution

Table below shows GSEA type pathway analysis on a best SNP per approach, genome wide

WTCCC Bipolar Disorder

Best p-value per gene

Pathway	Nominal P	FDR	
GO0007411	< 1e-3	0.003	Axon growth cone guidance
hsa04510	< 1e-3	0.005	Focal adhesion
hsa00040	< 1e-3	0.006	Pentose and glucuronate interconversions
GO0019198	< 1e-3	0.006	The catalysis of phosphate removal from a phosphotyrosine using aspartic acid as a nucleophile in a metal-dependent manner
GO0003779	< 1e-3	0.028	Membrane associated actin binding
hsa04512	< 1e-3	0.031	ECM-receptor interaction

Sklar et al.

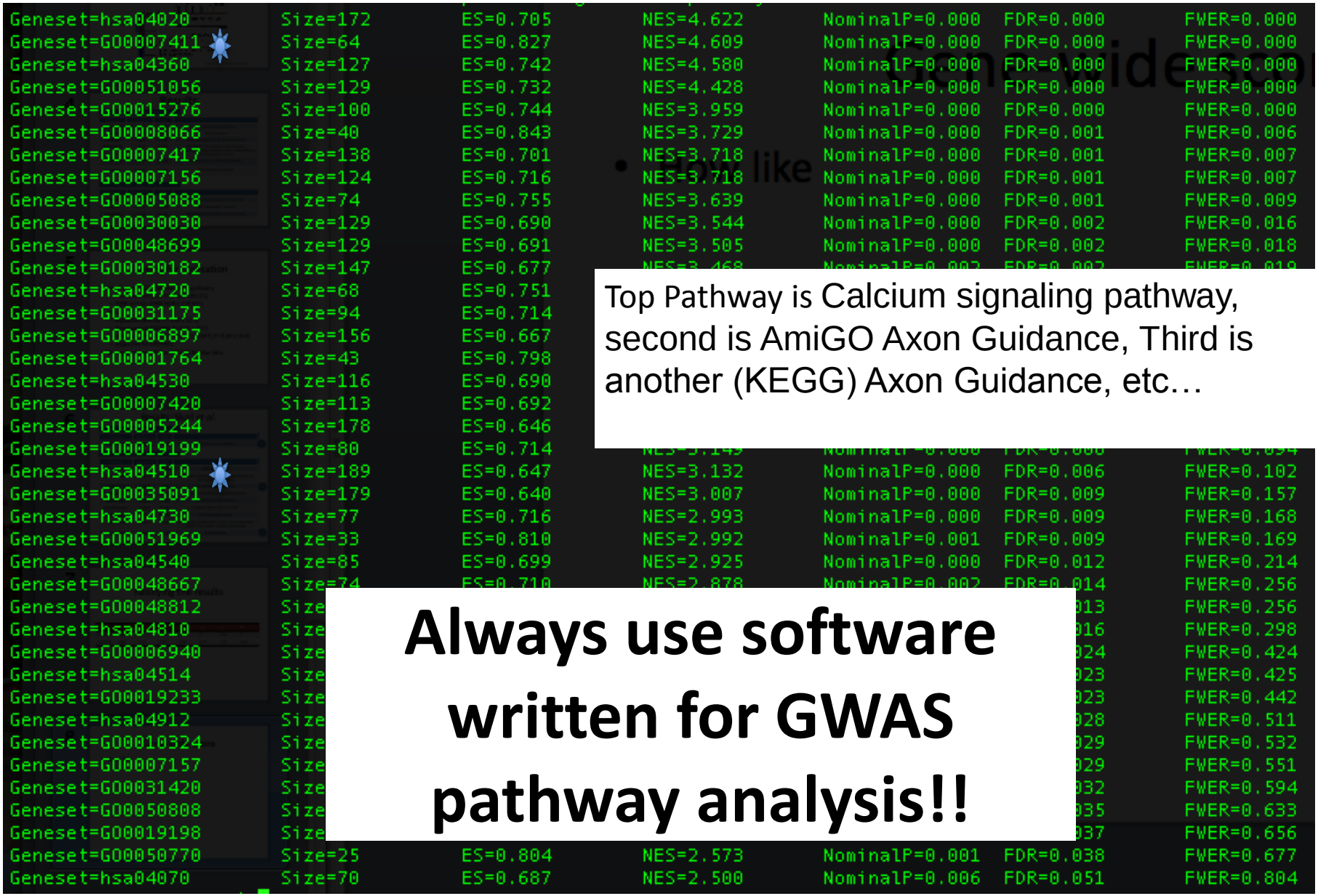
Pathway	Nominal P	FDR	
GO0007411	< 1e-3	0.025	Axon growth cone guidance

Genetics Falsifying the Results

	T1D	T2D	RA	HT	CAD	CD	BP
GO0007411	0.004	0	0.004	0.001	0	0.002	0
hsa04510	0.081	0.037	0.143	0.001	0.021	0.003	0.095

FDRs for these two pathways in all the WTCCC GWAS datasets

What happens if you make gene size the statistic?



Geneset	Size	ES	NES	NominalP	FDR	FWER
Geneset=hsa04020	Size=172	ES=0.705	NES=4.622	NominalP=0.000	FDR=0.000	FWER=0.000
Geneset=GO0007411	Size=64	ES=0.827	NES=4.609	NominalP=0.000	FDR=0.000	FWER=0.000
Geneset=hsa04360	Size=127	ES=0.742	NES=4.580	NominalP=0.000	FDR=0.000	FWER=0.000
Geneset=GO0051056	Size=129	ES=0.732	NES=4.428	NominalP=0.000	FDR=0.000	FWER=0.000
Geneset=GO0015276	Size=100	ES=0.744	NES=3.959	NominalP=0.000	FDR=0.000	FWER=0.000
Geneset=GO0008066	Size=40	ES=0.843	NES=3.729	NominalP=0.000	FDR=0.001	FWER=0.006
Geneset=GO0007417	Size=138	ES=0.701	NES=3.718	NominalP=0.000	FDR=0.001	FWER=0.007
Geneset=GO0007156	Size=124	ES=0.716	NES=3.718	NominalP=0.000	FDR=0.001	FWER=0.007
Geneset=GO0005088	Size=74	ES=0.755	NES=3.639	NominalP=0.000	FDR=0.001	FWER=0.009
Geneset=GO0030030	Size=129	ES=0.690	NES=3.544	NominalP=0.000	FDR=0.002	FWER=0.016
Geneset=GO0048699	Size=129	ES=0.691	NES=3.505	NominalP=0.000	FDR=0.002	FWER=0.018
Geneset=GO0030182	Size=147	ES=0.677				
Geneset=hsa04720	Size=68	ES=0.751				
Geneset=GO0031175	Size=94	ES=0.714				
Geneset=GO0006897	Size=156	ES=0.667				
Geneset=GO0001764	Size=43	ES=0.798				
Geneset=hsa04530	Size=116	ES=0.690				
Geneset=GO0007420	Size=113	ES=0.692				
Geneset=GO0005244	Size=178	ES=0.646				
Geneset=GO0019199	Size=80	ES=0.714				
Geneset=hsa04510	Size=189	ES=0.647	NES=3.132	NominalP=0.000	FDR=0.006	FWER=0.102
Geneset=GO0035091	Size=179	ES=0.640	NES=3.007	NominalP=0.000	FDR=0.009	FWER=0.157
Geneset=hsa04730	Size=77	ES=0.716	NES=2.993	NominalP=0.000	FDR=0.009	FWER=0.168
Geneset=GO0051969	Size=33	ES=0.810	NES=2.992	NominalP=0.001	FDR=0.009	FWER=0.169
Geneset=hsa04540	Size=85	ES=0.699	NES=2.925	NominalP=0.000	FDR=0.012	FWER=0.214
Geneset=GO0048667	Size=74	ES=0.710	NES=2.878	NominalP=0.002	FDR=0.014	FWER=0.256
Geneset=GO0048812	Size				013	FWER=0.256
Geneset=hsa04810	Size				016	FWER=0.298
Geneset=GO0006940	Size				024	FWER=0.424
Geneset=hsa04514	Size				023	FWER=0.425
Geneset=GO0019233	Size				023	FWER=0.442
Geneset=hsa04912	Size				028	FWER=0.511
Geneset=GO0010324	Size				029	FWER=0.532
Geneset=GO0007157	Size				029	FWER=0.551
Geneset=GO0031420	Size				032	FWER=0.594
Geneset=GO0050808	Size				035	FWER=0.633
Geneset=GO0019198	Size				037	FWER=0.656
Geneset=GO0050770	Size=25	ES=0.804	NES=2.573	NominalP=0.001	FDR=0.038	FWER=0.677
Geneset=hsa04070	Size=70	ES=0.687	NES=2.500	NominalP=0.006	FDR=0.051	FWER=0.804

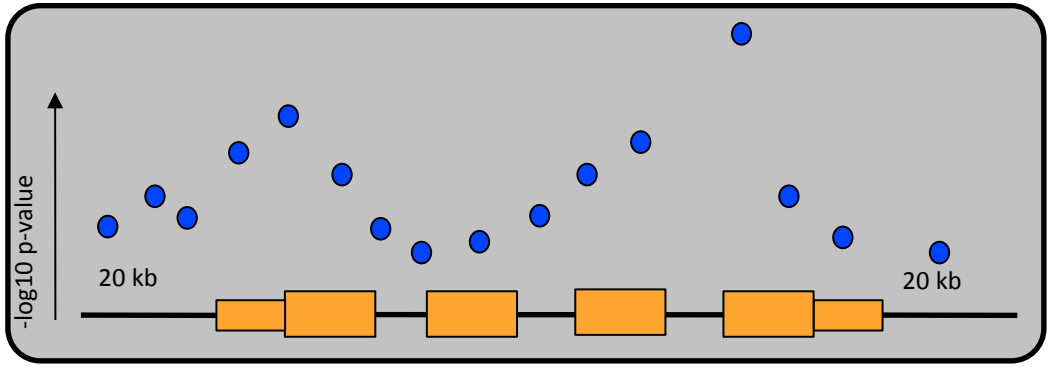
Top Pathway is Calcium signaling pathway, second is AmiGO Axon Guidance, Third is another (KEGG) Axon Guidance, etc…

Always use software written for GWAS pathway analysis!!

Two Main routes.

Gene-wise first

- Derive a gene wide statistic.
- Then assess gene sets using many possible methods.



- ✴ Correct min p-value (e.g. Sidak)
- ✴ Threshold
- ✴ Combine p-values (e.g. Fisher's method)

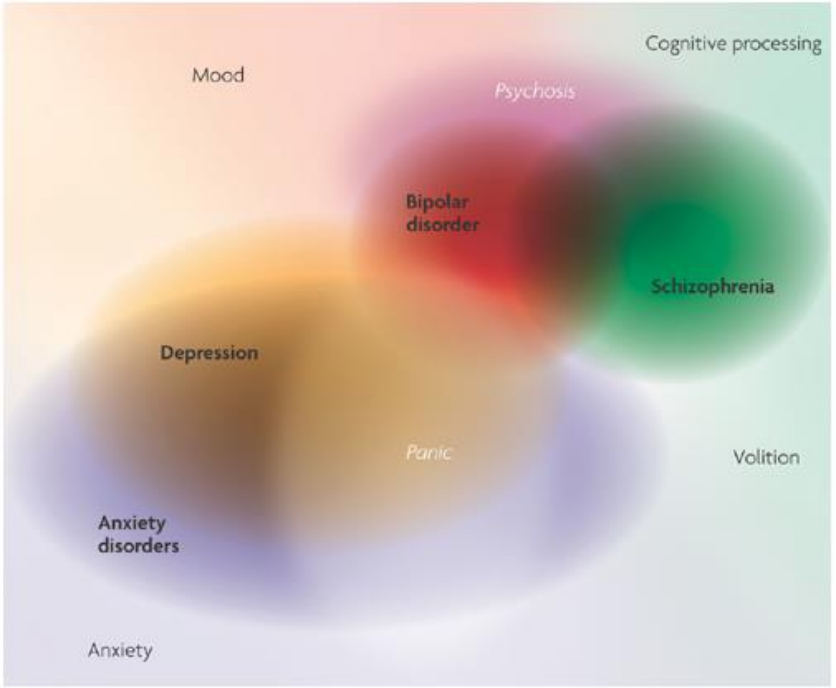
Direct Gene-set analyses

- Treat a pathway as one large gene.

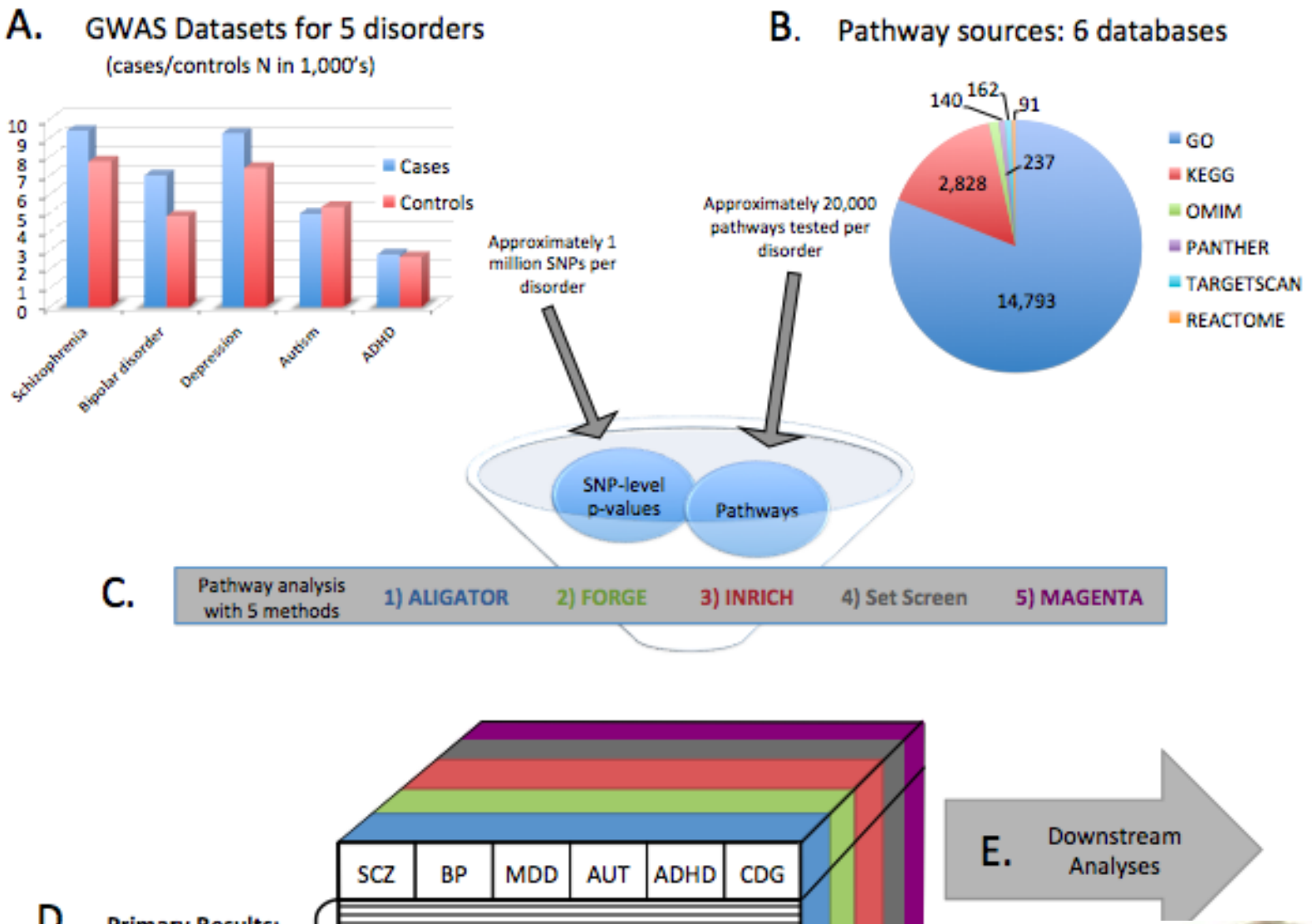
I. Pedroso, G. Breen, in Gene Set Analysis and Network Analysis for Genome-Wide Association Studies, A. Al-Chalabi, L. Almasy, Eds. (Cold Spring Harbor Laboratory Press, 2009)

PGC CDG GWAS Pathway Analysis

- 5 methods
- 5 diseases
- Schizophrenia
- Manic Depression
- Major Depression
- Autism
- ADHD



Nature Reviews | Genetics



Laramie Duncan

Methods

- The combined GWAS dataset of the five disorders comprised 61,220 case and controls.
- We are using pathway methods on these data.
- For all methods, we used a consensus set of pathways from a range of sources and a consensus set of genes with defined boundaries.
- We compared the intersections of the results of both pathways and of pathways across diseases.

Samples

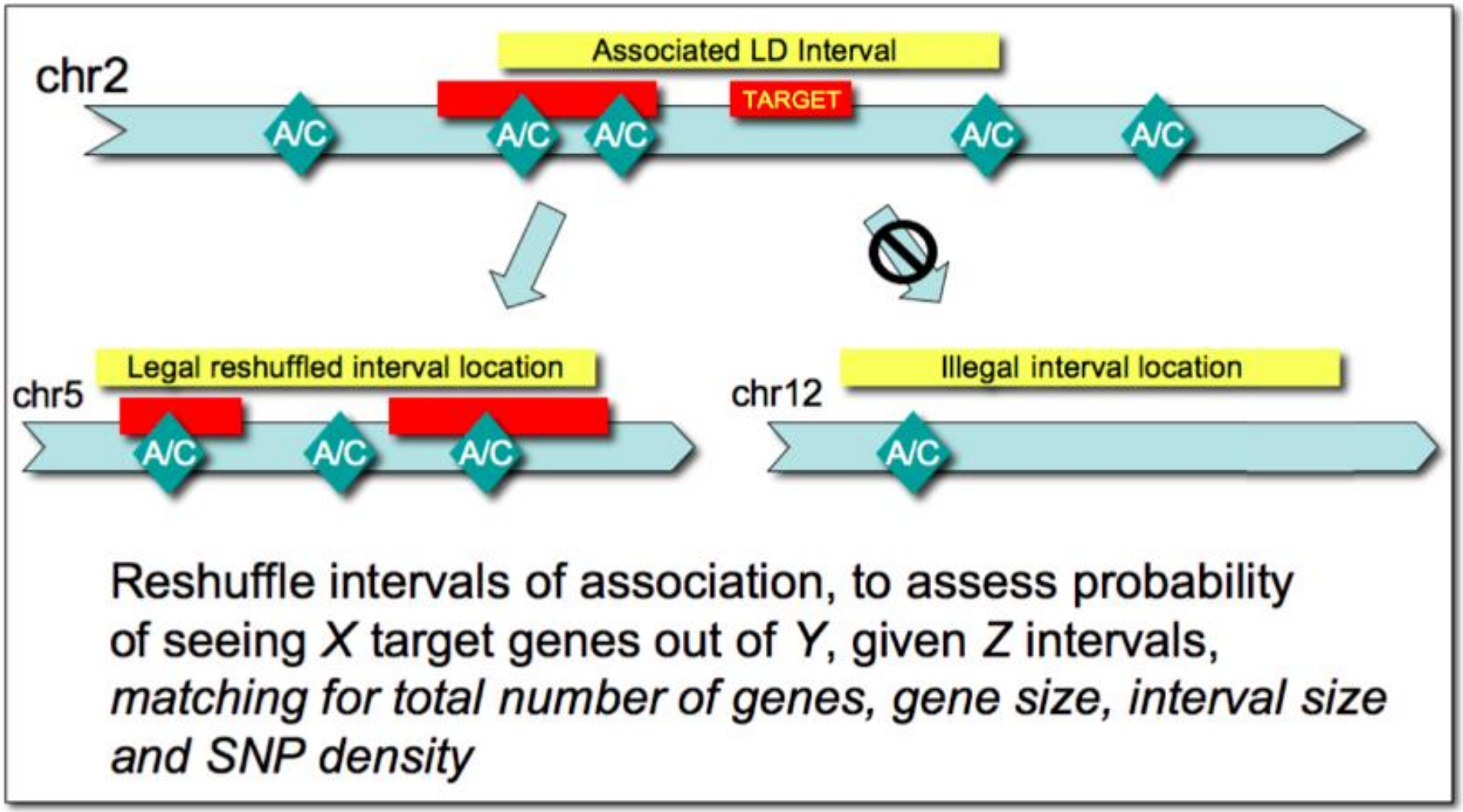
- The combined GWAS dataset of the five disorders comprised 60K case and controls.
- Major Depression (9,227 / 7,383)
- Manic Depression (Bipolar Disorder) (6,990/ 4,820)
- Schizophrenia (9,370 / 7,736)
- Autism (4,949 / 5,314) Trios
- Attention Deficit Hyperactivity Disorder (ADHD) (2,787/ 2,635) Trios

Pathway Analysis Methods

- Thresholded best/number in gene/region methods
  - ALIGATOR
  - INRICH
  - MAGENTA
- Gene-wide/Pathway-wide methods
  - Set-screen test
  - FORGE

INRICH: *interval enrichment*

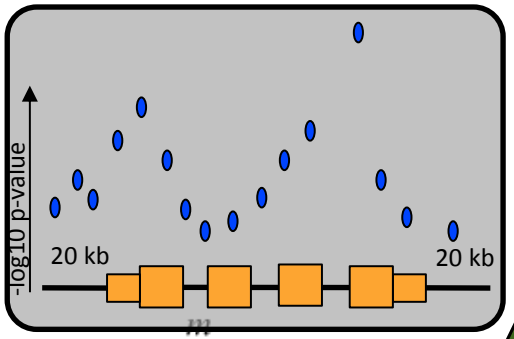
"Do we see more associated genes in set X compared to chance?"



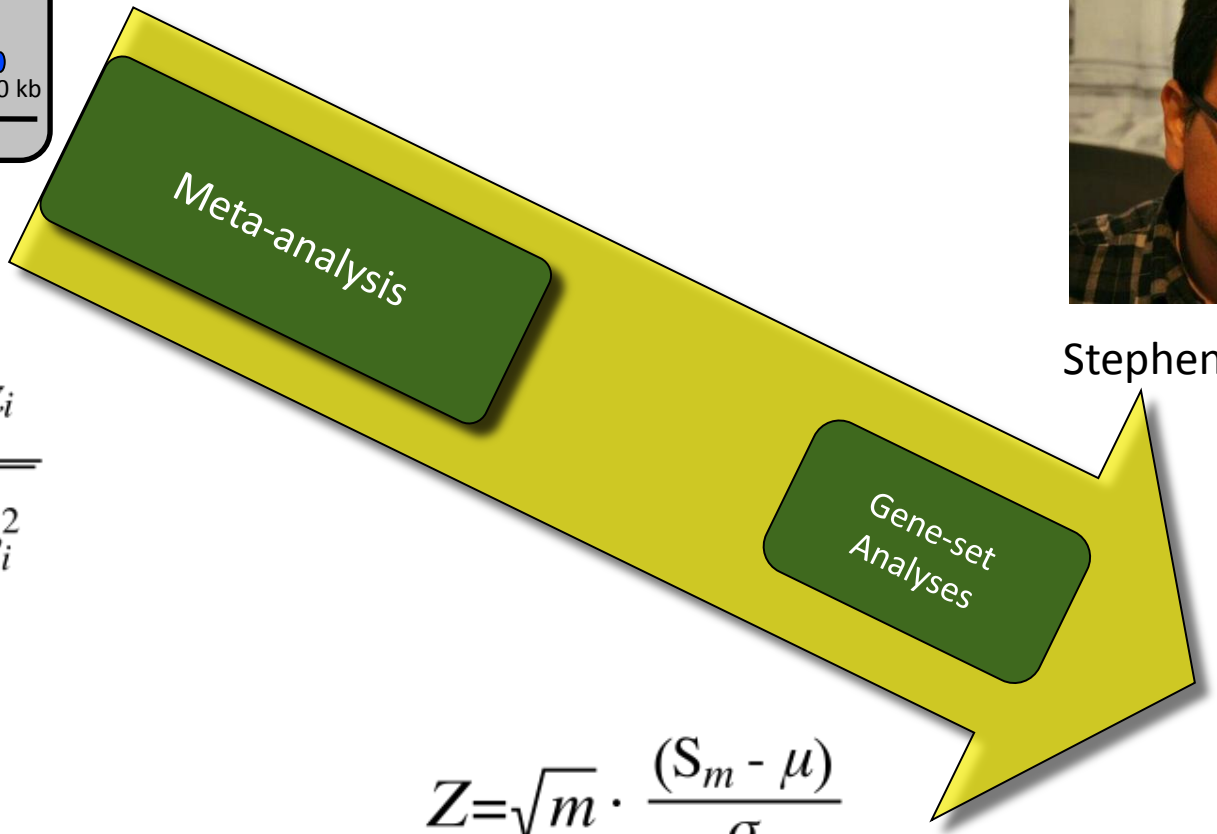
Reshuffle intervals of association, to assess probability of seeing *X* target genes out of *Y*, given *Z* intervals, *matching for total number of genes, gene size, interval size and SNP density*

FORGE: gene p-values by LD corrected metaanalysis and GSA

Pedroso….Breen et al., Biol Psych 2012



$$M_F = -2\sum_{i=1}^{m} \ln(p_i)$$



$$Z = \frac{\sum_{i=1}^{k} w_i z_i}{\sqrt{\sum_{i=1}^{k} w_i^2}}$$

$$Z = \sqrt{m} \cdot \frac{(S_m - \mu)}{\sigma}$$

Stephen Newhouse

Pre-defined pathway databases

- Public databases
  - KEGG
  - GO
  - Reactome
  - PANTHER
  - OMIM
  - Targetscan
- Custom databases available to the group
  - Synaptic gene database (Danielle Posthuma)

Details on Databases

Database	Method	Details
KEGG	Expert curation	Metabolism Genetic Information Processing Environmental Information Processing Cellular Processes Organismal Systems Human Diseases Drug Development
GO	Expert curation	Controlled vocabulary: associated biological processes, cellular components and molecular functions
Reactome	Expert curation	Reactions Pathways
PANTHER	Expert curation	Primarily signaling pathways
OMIM	Repository	All known Mendelian disorders
TargetScan	Computational prediction	Computational prediction of microRNA target genes based on context score, which is a function of site-type, 3' pairing, local AU content, and position

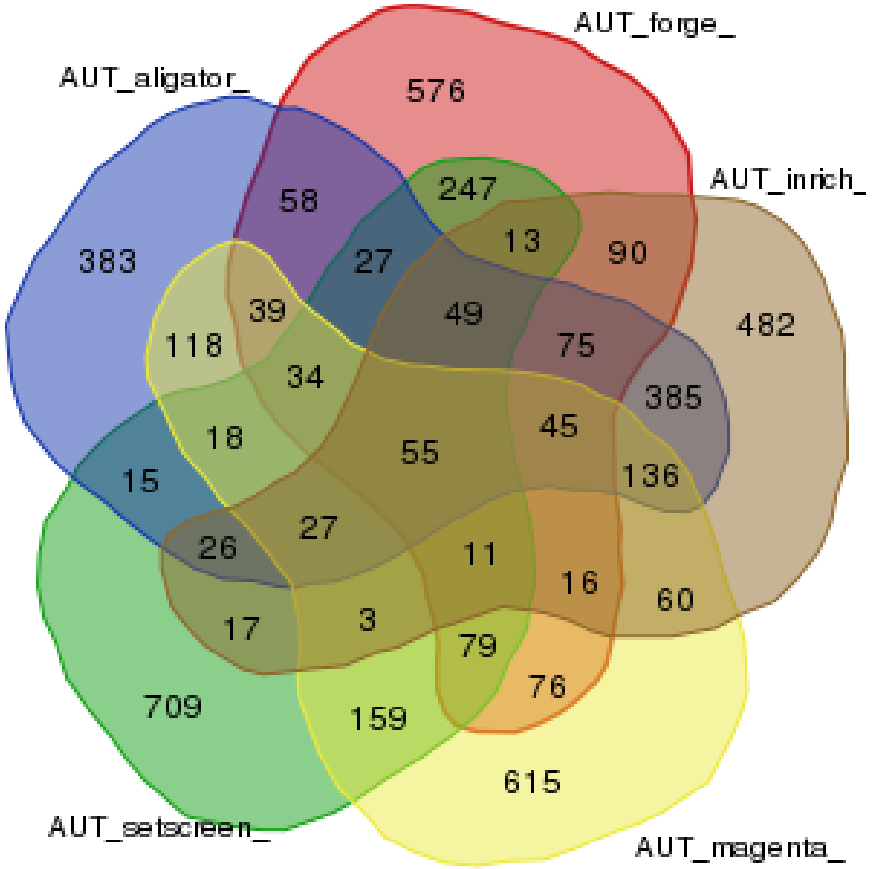
Databases – lots of them and lots of work

Colm O`Dushlaine
Lizzy Rossin

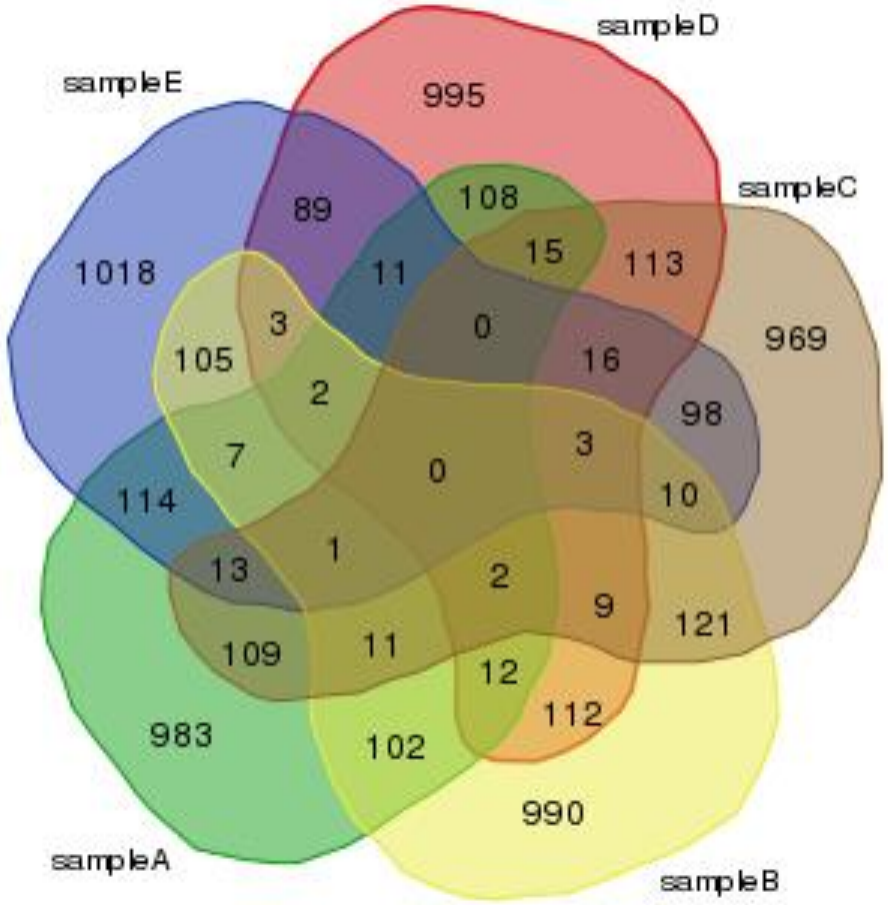
Database	# Genes covered	# Pathways	Median Pathway Size (min-max)	Link to Histogram
KEGG	5952	232	52 (1-1131)	Pathway Size - KEGG.pdf
GO	8589	7112	2 (1-2407)	Pathway Size - GO.pdf
Reactome	4539 5077	3526 (Reaction) 1086 (Pathway)	3 (1-434) 14 (1-934)	Reaction Size - Reactome.pdf
PANTHER	2170	140	16 (1-287)	Pathway Size - PANTHER.pdf
OMIM	6983	4712	2 (1-22)	Pathway Size - OMIM.pdf
TargetScan	11095	162	173 (1-1240)	Pathway Size - TargetScan.pdf

Methods Comparison

- What is the overlap in the top 10% of pathways?



OBSERVED



EXPECTED

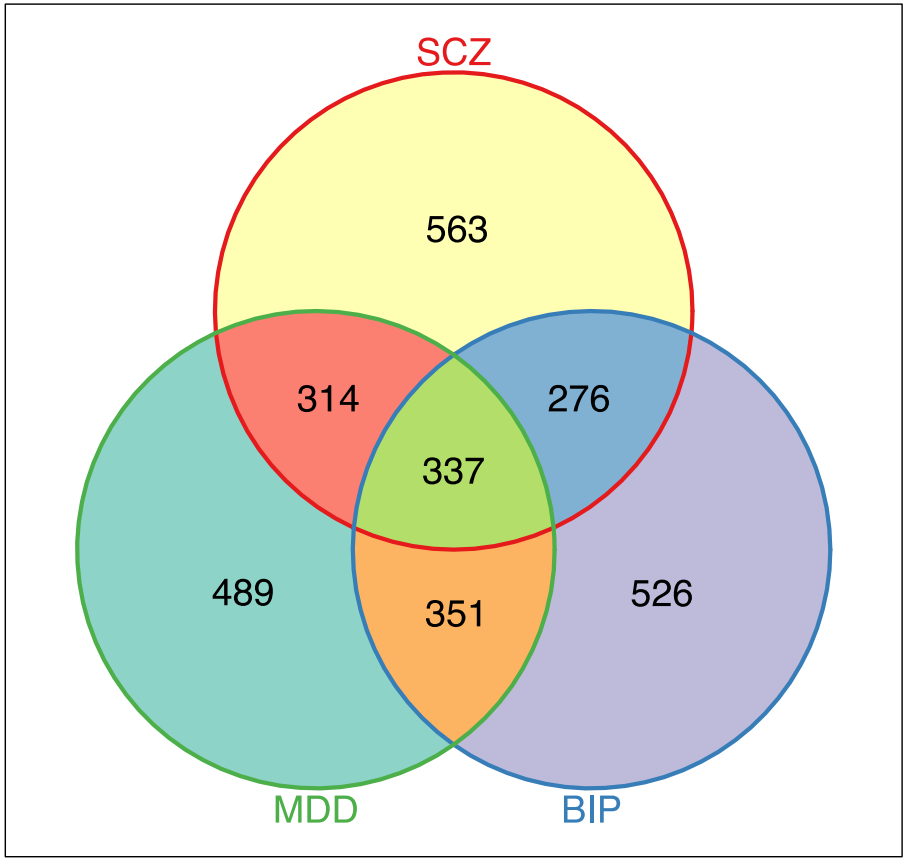
Comparing methods and disorders

- Each method has its advantages.
- Underlying genetic architecture.
- Decided to go down CNV route.
- **Comparing Disorders:**
- BIPOLAR, MDD, SCZ
- Autism and ADD (PGC1 versions) are less well powered, making comparisons difficult

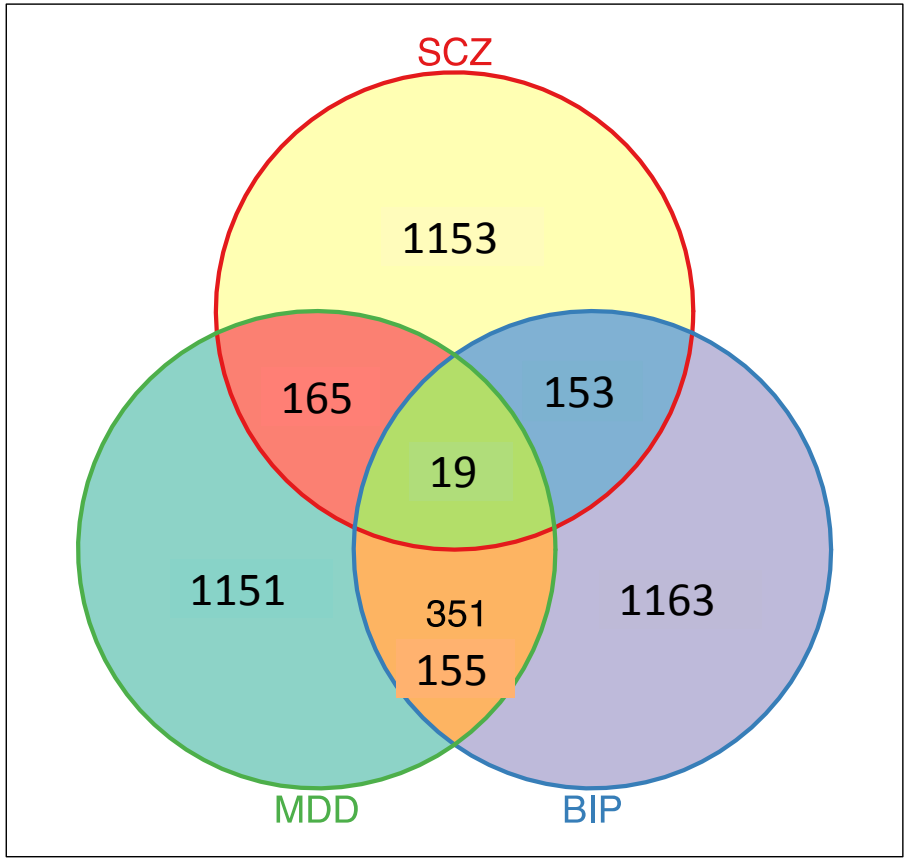
Phil Lee

PGC Disorders

- SCZ, BIPOLAR and MDD are well-powered
- Overlap in the top 10% of pathways? (INRICH)

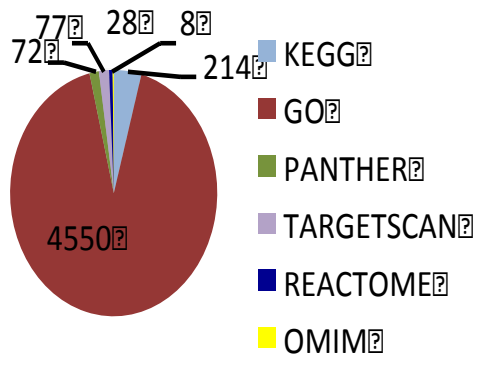


OBSERVED



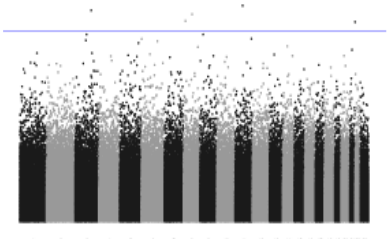
EXPECTED

A



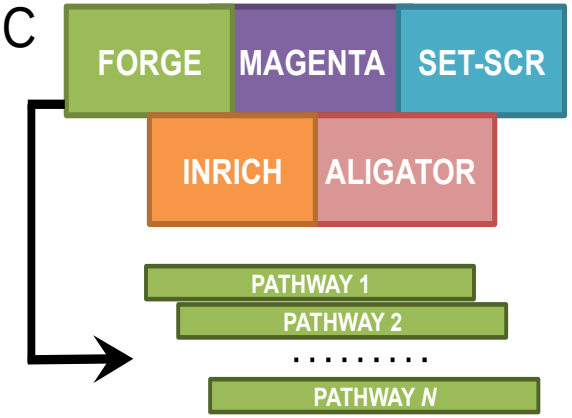
Step 1: Gather pathways from public databases, prune redundancy.

B



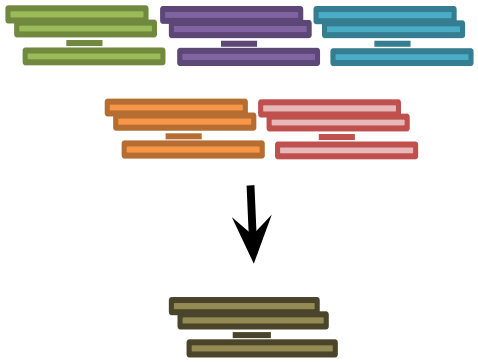
Step 2: Take raw p-values from GWAS

C



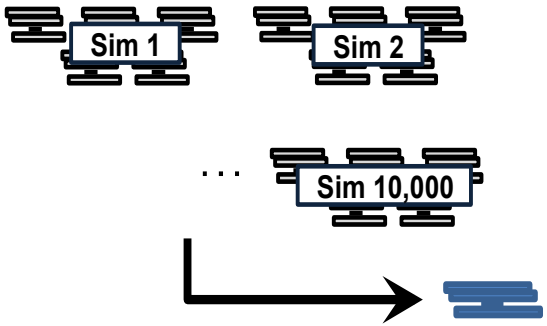
Step 3: Score each pathway using 5 methods

D



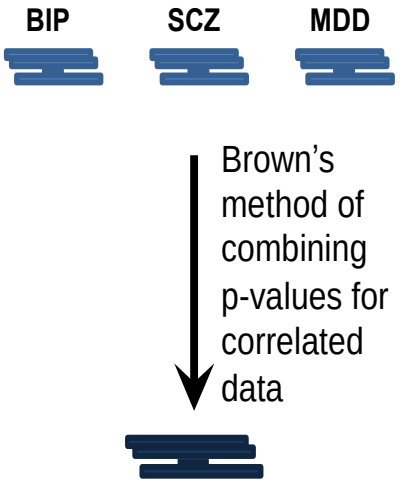
Step 4: Rank pathways and obtain an average rank across methods

E **Step 5a:** Generate 10,000 null simulated datasets, respecting the correlations between methods



Step 5b: Calculate empirical p-value by comparing average ranks in disease data to average ranks in simulated data

F



Step 6: Derive a combined p-value for three psychiatric disorders to highlight common etiologic mechanisms.

Datasets

- PGC-CDG datasets
- Dataset are independent
- Null dataset (simulated from HapMap)
- SNPs with MAF>0.01 and info >0.8
- Analysis restricted to protein-coding genes
- Window around genes: 35-10kb (eQTL RNAseq)
- Analysis without MHC

Peter Holmans

Obtaining significance of enrichments (1)

- Rank pathways by enrichment p-value for each method
- Measure of enrichment: average rank
- Calculate correlations of p-values for each pair of methods using null dataset
- Generate 1,000,000 sets of 6 uniform variates with these pairwise correlations
- Generate a replicate “study” by sampling one of these sets for each pathway (corresponding to enrichment p-values)

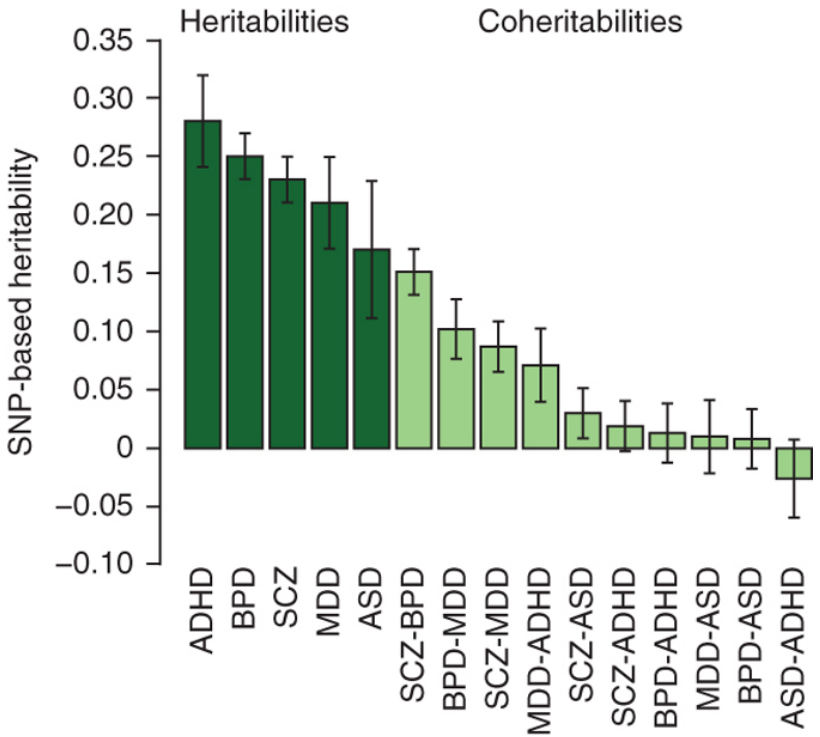
Peter Holmans

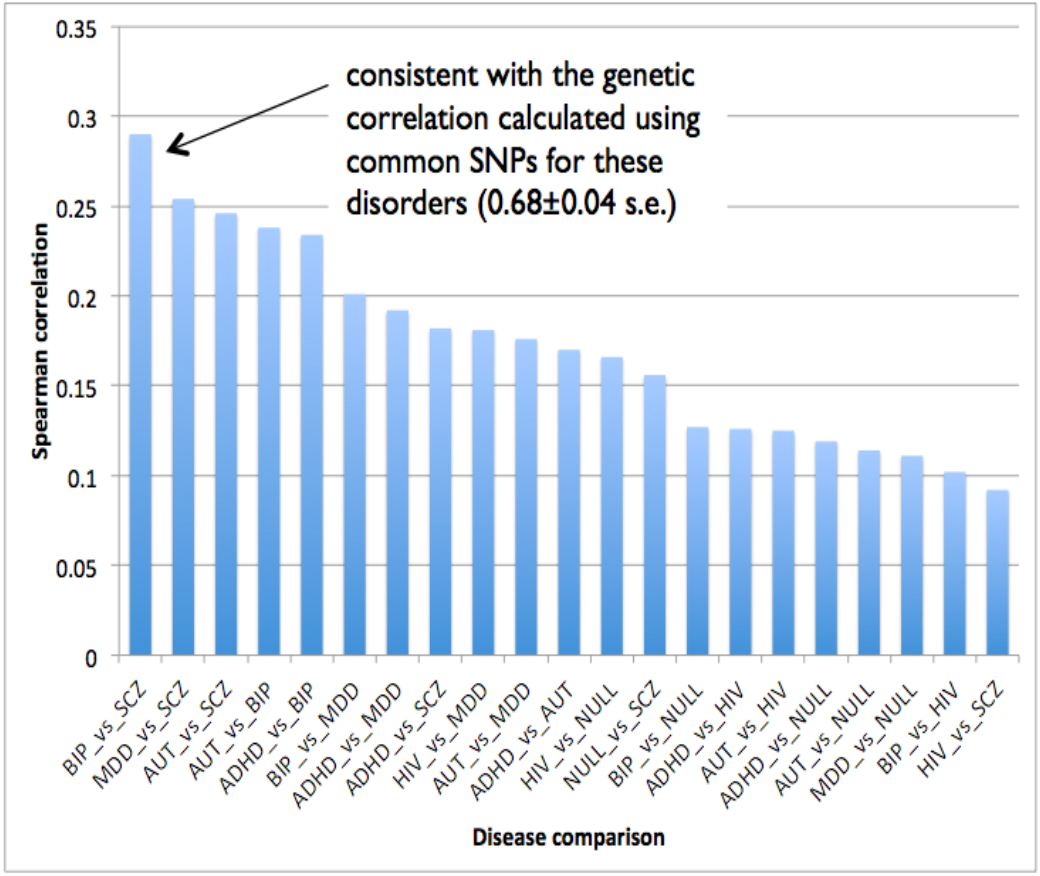
Obtaining significance of enrichments (2)

- Rank pathways by enrichment p-value within each method
- Compare average rank to that in the real data
- Repeat 1,000,000 times to obtain pathway-specific enrichment p-value for each pathway
- Correction for multiple testing of pathways: q-value

Peter Holmans

Pathway correlations between disordered resemble GREML bivariate analyses.



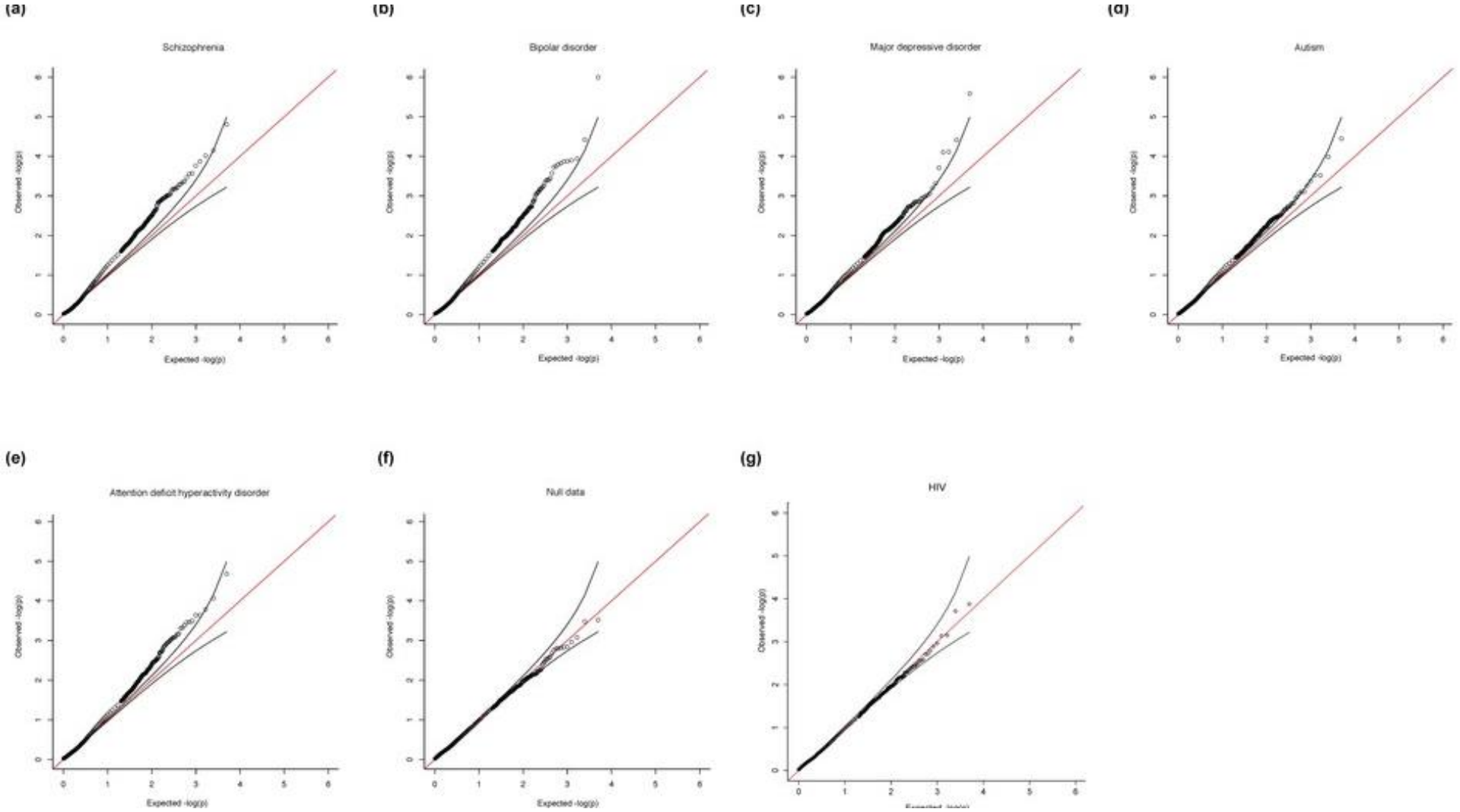


Top Pathways

- Results from a combined analysis of 5 methods on the 5 PGC1 CDG datasets but we focused on the three adult disorders.
- Overlapping controls and cases removed.
- The following slides list the top pathways by FDR Q-value.
- Thus, the results reflect the average across the 5 methods used.
- The results follow the typical PGC pattern but are more exciting and positive than we expected.
- We defined significant q-values as <0.05 and suggestive <0.5.
- We excluded the MHC region (high-LD genes also in that region on chromosome 6)

QQ plots for each disorder and for HIV and Null sets

Order is (a) SCZ, (b) BIP, (c) MDD, (d) AUT, (e) ADD, (f) null data from permutations and, (g) HIV.



Meta-analysis of the pathway results SCZ, BIP, MDD
16 pathways with a FDR <0.05, 49 with an FDR<0.1.

# methods	Av. rank	p rank	q-value	Pathway ID	Description
BIP					
5	17	1.01E-06	0.005	GO:51568	histone H3-K4 methylation
5	50.4	3.82E-05	0.093	path:hsa05218	Melanoma
5	79.2	1.16E-04	0.093	GO:7129	(chromosomal) synapsis
5	81.8	1.27E-04	0.093	path:hsa05213	Endometrial cancer
5	83.3	1.34E-04	0.093	P00003	Alzheimer_disease-amyloid_secretase_pathway
5	83.4	1.35E-04	0.093	path:hsa05215	Prostate cancer
5	87	1.50E-04	0.093	path:hsa05216	Thyroid cancer
4	89.5	1.59E-04	0.093	GO:90066	regulation of anatomical structure size
5	95.6	1.81E-04	0.093	path:hsa05214	Glioma
5	96.9	1.87E-04	0.093	GO:70192	chromosome organization involved in meiosis
SCZ					
5	38.4	1.58E-05	0.078	GO:14069	postsynaptic density
5	68.6	7.15E-05	0.160	GO:45211	postsynaptic membrane
5	76.8	9.67E-05	0.160	GO:43197	dendritic spine
5	85.4	1.36E-04	0.168	GO:51568	histone H3-K4 methylation
5	95.8	1.74E-04	0.173	GO:33267	axon part
MDD					
5	25.4	2.63E-06	0.012	GO:8601	protein phosphatase type 2A regulator activity
5	54.6	3.88E-05	0.092	GO:34330	cell junction organization
5	68.8	7.70E-05	0.094	GO:43297	apical junction assembly
5	70	7.92E-05	0.094	GO:45216	cell-cell junction organization
5	99.8	1.97E-04	0.186	GO:31056	regulation of histone modification

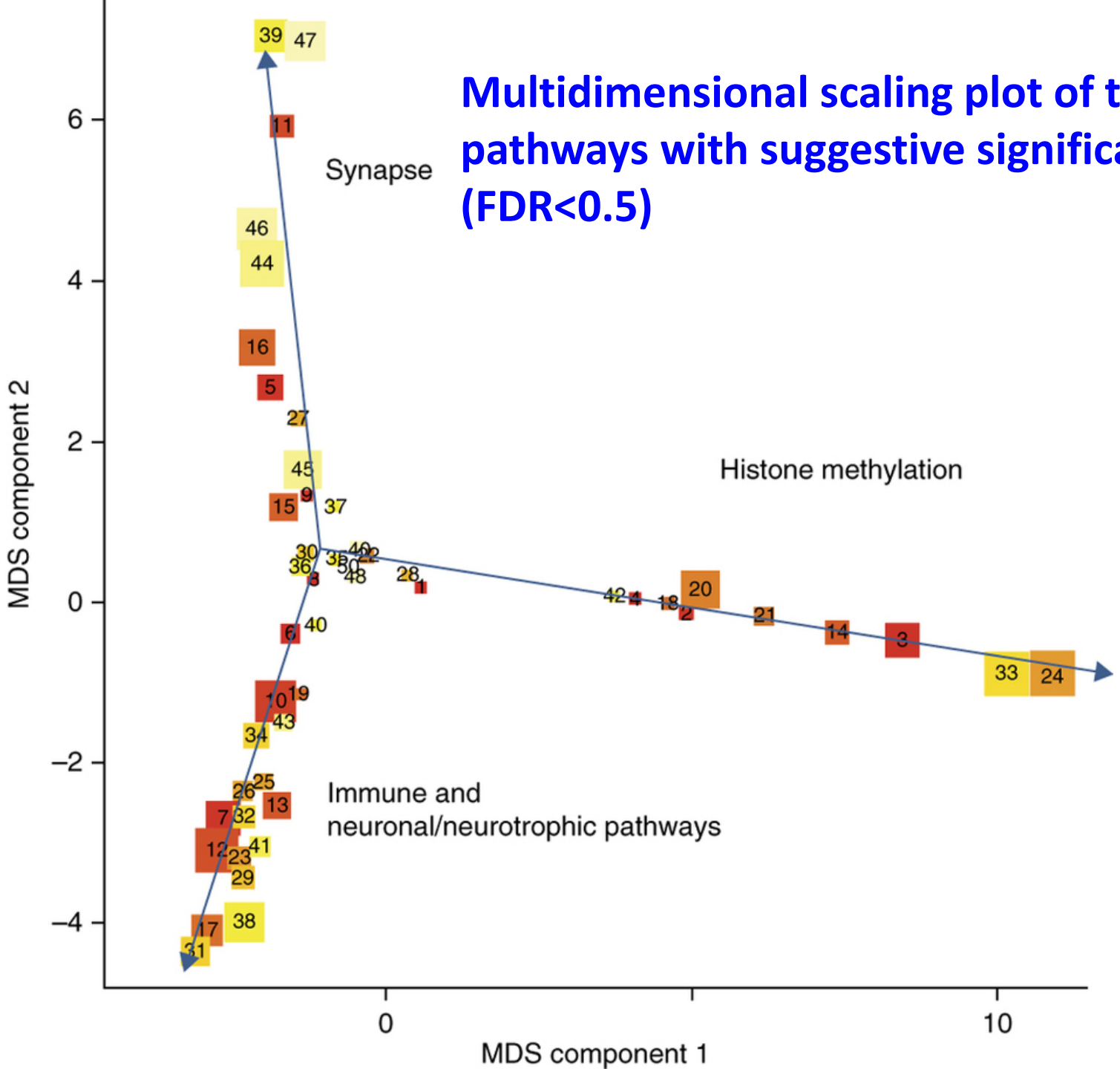
What are key difference in meta-analysis of Pathway Results?

- Each disorder gave promising but not statistically compelling evidence for pathway association.
- Analyse each disorder's pathways and then combine and meta-analyse.
- May be much more powerful than SNP meta analysis.
- Robust to
  - Allelic heterogeneity within GENES and within PATHWAYS across diseases.
  - Allows for a multitude of weaker effects.
  - Modulation of the pathways can differ across disorders.

Meta-analysis of the pathway results SCZ, BIP, MDD
16 pathways with a FDR <0.05, 49 with an FDR<0.1.

Pathway Level Meta Analysis across three disorders.

BIP	MDD	SCZ	Combined P	Q Value	Pathway ID	Description
0	0.0592	0.0001	5.75E-08	0.0003	GO:51568	histone H3-K4 methylation
0.0004	0.05	0.0006	1.46E-05	0.0362	GO:16571	histone methylation
0.0004	0.1462	0.0011	4.73E-05	0.0414	GO:43414	macromolecule methylation
0.0008	0.063	0.0014	5.10E-05	0.0414	GO:34968	histone lysine methylation
0.42	0.0001	0.0023	5.58E-05	0.0414	GO:45216	cell-cell junction organization
0.0001	0.091	0.0064	5.69E-05	0.0414	P00003	Alzheimer_disease-amyloid_secretase_pathway
0.0007	0.0495	0.0024	5.86E-05	0.0414	P04393	Ras_Pathway
0.312	0	0.1286	7.12E-05	0.0422	GO:8601	protein phosphatase type 2A regulato activity
0.898	0.0001	0.0017	7.83E-05	0.0422	GO:43297	apical junction assembly
0.0013	0.0207	0.0055	9.25E-05	0.0422	P00052	TGF-beta_signaling_pathway
0.489	0.0203	0	9.53E-05	0.0422	GO:14069	postsynaptic density
0.0085	0.0009	0.0239	0.0001	0.0422	GO:32869	cellular response to insulin stimulus
0.0188	0.0054	0.0022	0.0001	0.045	P00010	B_cell_activation
0.0023	0.2988	0.0003	0.0001	0.045	GO:8757	S-adenosylmethionine-dependent methyltransferase activity
0.0073	0.008	0.0044	0.0001	0.0454	GO:23061	signal release
0.459	0	0.0168	0.0002	0.0473	GO:34330	cell junction organization


Multidimensional scaling plot of top 50 pathways with suggestive significance (FDR<0.5)
Synapse
Histone methylation
Immune and
neuronal/neurotrophic pathways
MDS component 1
MDS component 2
6
4
2
0
−2
−4
0
10

Chromatin and histones

- Chromatin is the combination of DNA and proteins that make up the contents of the nucleus of a cell.
- The primary protein components of chromatin are histones that compact the DNA.
- Five major families of histones exist: H1/H5, H2A, H2B, H3 and H4.
- Histone methylation is a process by which methyl groups are transferred to amino acids of histone proteins of chromosomes.
- Methylation can modify histones so that different portions of chromatin are activated or inactivated – turning genes in DNA "off" and "on”.

Histone Methylation

- Replicated environmental risks for SCZ occur in 1st trimester when the epigenome is known to be particularly labile.
- Rapid cell-replication is occurring and the standard epigenetic signals, including histone H3-K4 and lysine methylation are being laid down.
- Dysregulation in histone methylation pathways is indicated as a key common etiological mechanism for adult psychiatric disorders.
- Animal studies where prefrontal dysfunction and defects in learning and memory are induced in a gene knockout of a histone methyltransferase *MLL1.*

In the Genome

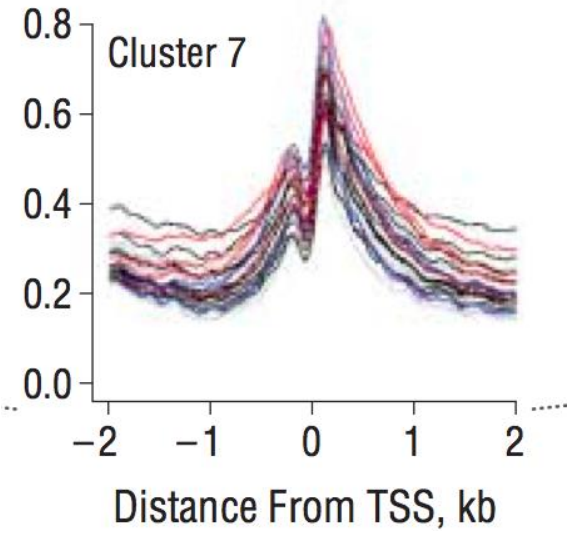


Figure 2: DNA methylation of CpG islands is mutually exclusive with H3K4 methylation in sperm.

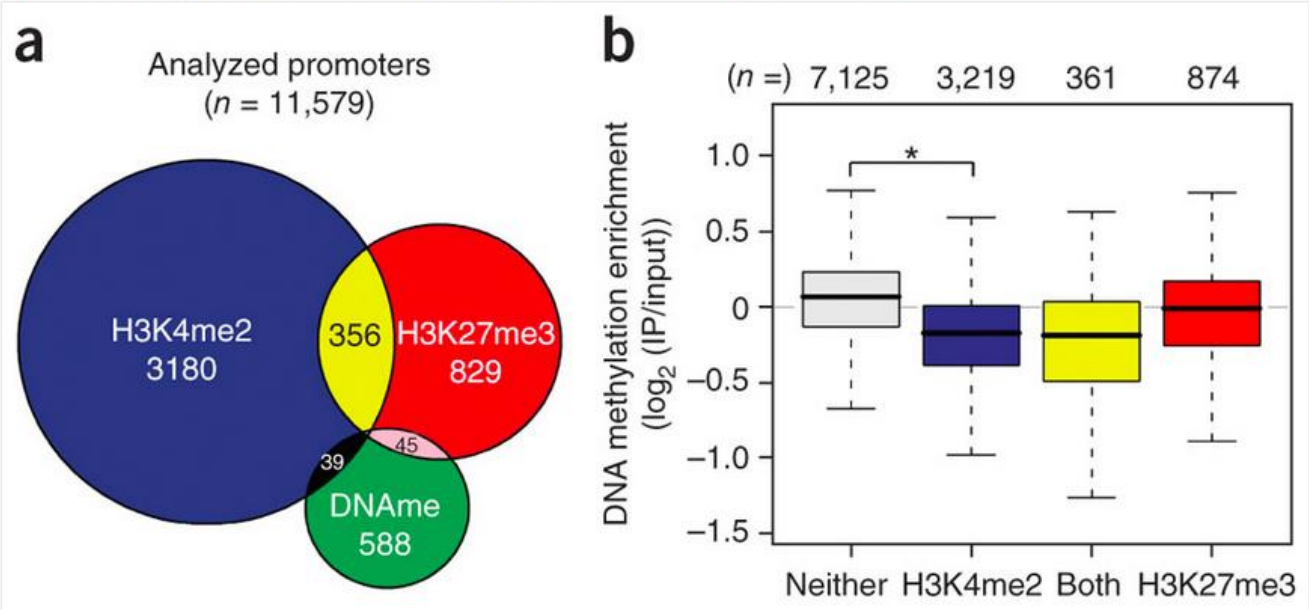
From

Repressive and active histone methylation mark distinct promoters in human and mouse spermatozoa

Urszula Brykczynska, Mizue Hisano, Serap Erkek, Liliana Ramos, Edward J Oakeley, Tim C Roloff, Christian Beisel, Dirk Schübeler, Michael B Stadler & Antoine H F M Peters

Nature Structural & Molecular Biology **17**, 679–687 (2010) | doi:10.1038/nsmb.1821

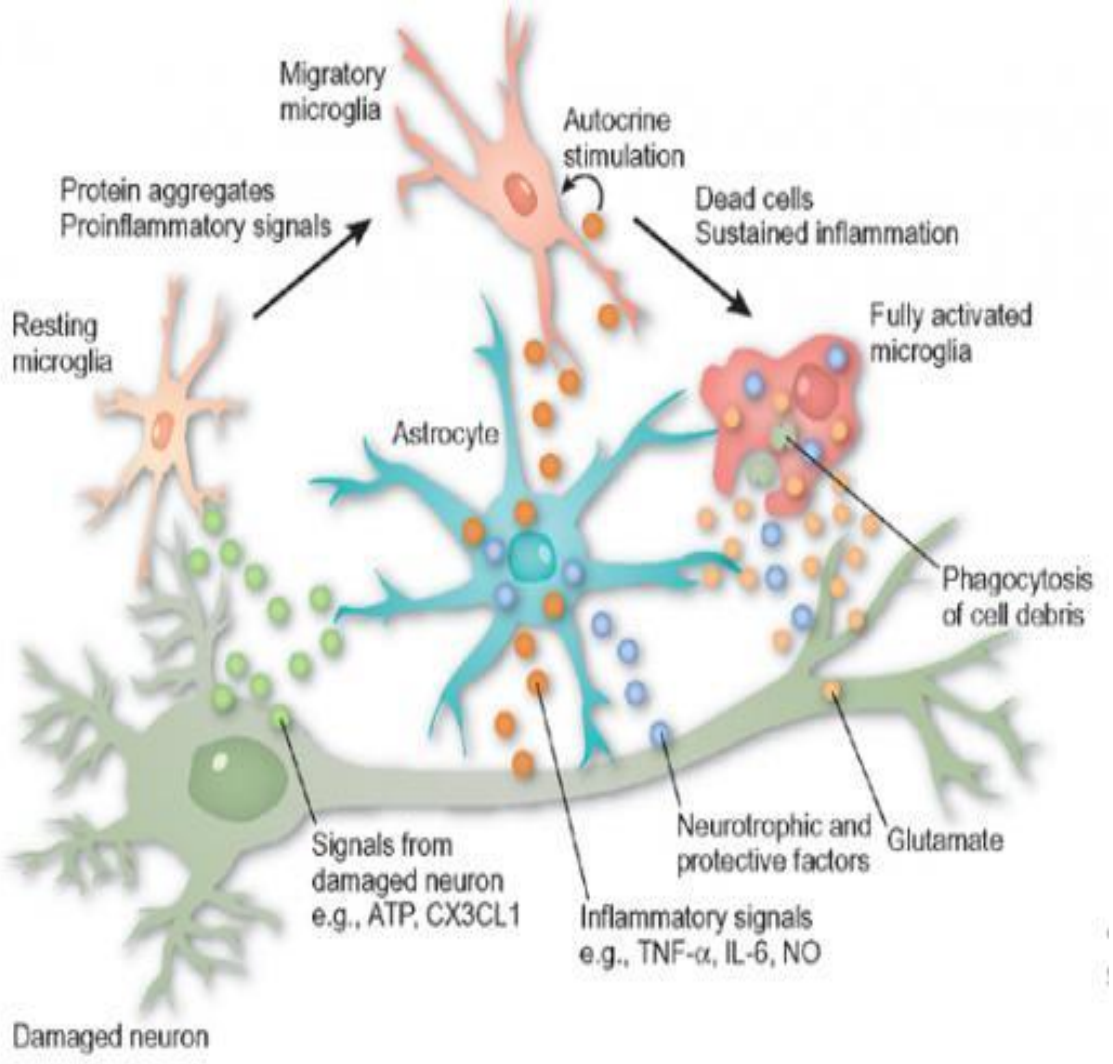
Figure 2: DNA methylation of CpG islands is mutually exclusive with H3K4 methylation in sperm.



Chrom	Start	End	Name
chr7	104,654,636	104,754,531	**MLL5**
chr7	151,832,009	152,133,090	**MLL3**
chr16	30,968,614	30,995,985	**SETD1A**
chr19	36,208,920	36,229,458	**AD000671.1**
chr1	205,055,269	205,091,143	**RBBP5**
chr11	118,307,227	118,395,411	**MLL**
chr8	37,963,010	37,997,225	**ASH2L**
chr12	122,242,637	122,270,562	**SETD1B**
chrX	44,732,422	44,971,843	**KDM6A**
chr18	47,808,712	47,814,674	**CXXC1**
chr7	154,735,519	154,794,621	**PAXIP1**
chr2	32,249,170	32,264,874	**DPY30**
chr3	52,288,436	52,312,659	**WDR82**
chr12	49,412,757	49,449,107	**MLL2**
chr9	137,001,209	137,025,090	**WDR5**

Immune Pathways

- Immense literature and GWAS for MHC region but we excluded it.
  - (a) genomic history and LD & (b) are other immune processes involved?
- Field clouded by somewhat controversial studies and concerns that finding confounded by/downstream of severe adverse environment.
- Genetic indications cleaner evidence overall – trait not state.
- Notable basic immunological pathways here:
  - *TGF-beta_signaling* (*P00052*)
  - *B_cell_activation* (*P00010)*
  - *T_cell_activation (P00053)*
- KEGG Infectious Disease pathways:
  - *Tuberculosis (hsa05152)*
  - *Hepatitis C (hsa05160)*
  - *Chagas disease (American trypanosomiasis, hsa05142).*



Immune pathways/components, Complement etc. have essential housekeeping and barber/hairdressers roles in the brain.

Makes for good stories:

- INFα treatment causes acute depression in about a third of recipients .
- Both HepC infection & INFα treatment associate with a range of additional neuropsychiatric symptoms.
- Chagas diseases cause stroke but phenothiazines such as chloropromazine are effective treatments in animal studies.

Effects on gene expression

- Examine whether top pathway genes that are associated with BIP+MDD+SCZ are enriched with brain expression regulatory SNPs (e.g., eQTLs, mQTLs)?
- Brain Co-expression networks

Top 10 pathway: genes

Pathway ID	Description	Ensembl[a] Gene #
GO:51568	histone H3-K4 methylation	15
GO:16571	histone methylation	55
P04393	Ras_Pathway	68
P00003	Alzheimer_disease-amyloid_secretase_pathway	66
GO:45216	cell-cell junction organization	111
P00052	TGF-beta_signaling_pathway	90
GO:14069	postsynaptic density	100
GO:34968	histone lysine methylation	44
P00010	B_cell_activation	59
GO:43414	macromolecule methylation	132
Top.10.path	**Top 10 pathways**	**556**

Brain eQTL/mQTL datasets

Src	Ref	Brain Region	Subject #	Expression	Genotype	1KG imputed
Meta Brain cis-eQTL (*In prep*)	Myers 2007	Cortex[a]	193	GSE8919	Affy 500k	yes
	Gibbs 2010	FTCX, TCTX	150	GSE15745	Illumina HapMap550	yes
	Webster 2009	Cortex[a]	176	GSE15222	Affy 500k	yes
	Colantuoni 2011	DPFC	112	GSE30272	Illumina 1M; HumanHap650	yes
	2,973 cis eQTL LD blocks with p < 1e-06 (kb=250; r2=0.4; clump-p1=1e-06; clump-p2=1e-04)					
UK Neurochip cis/trans eQTL	michael.weale@kcl.ac.uk; 1KG imputed; 134 subjects; 10 brain regions					
	20,616 cis/trans eQTL LD blocks with FDR 1% (kb=250; r2=0.4)					

Enrichment tests (Phil Lee)

For each of top pathways,

1) Extract LD-independent SNPs located in top 10 pathway genes (35/10K) with cdg3 meta-*p* < 5e-03

2) Count the # of brain eQTL LD-blocks overlapping with the pathway SNPs selected from (1).

3) Randomly select the same number of LD-independent SNPs associated with genic regions (N=20,007; 35/10K boundary) that satisfies the association significance (cdg3-*p*<5e-03) and matching MAF

4) Repeat (3) 100K times to assess the significance of the overlap from (2).

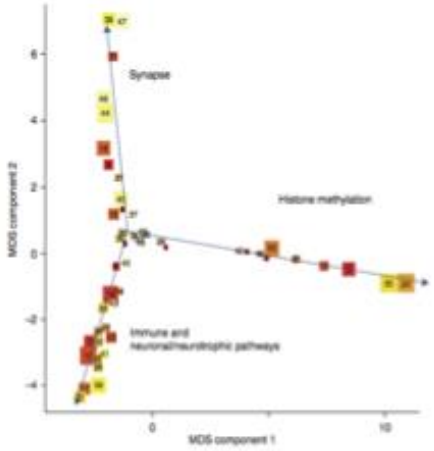
5) If significant,
    1) Replicate with the UK NeuroChip eQTL data
    2) Investigate non-brain disorders as negative examples

Enrichment testing results

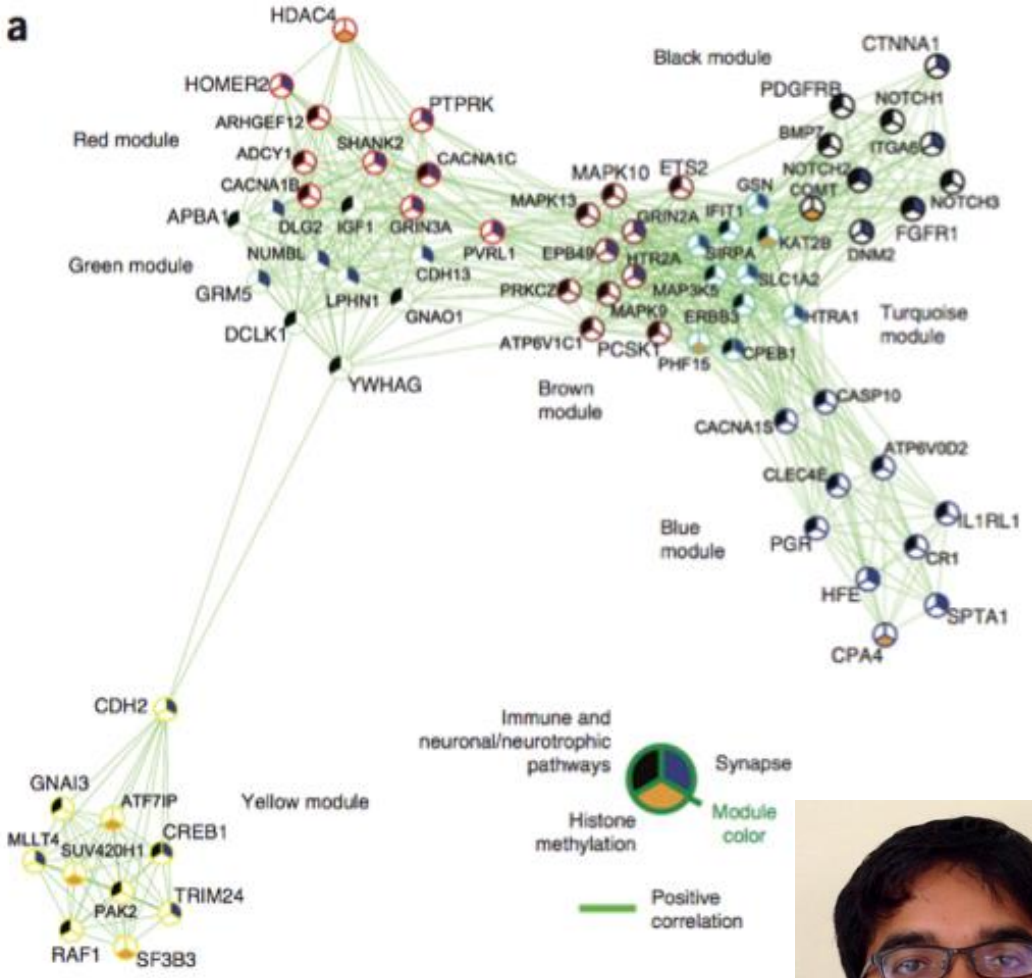
BIP/MDD/SCZ-associated methylation pathway genes are enriched with brain eQTLs, which suggests that their common disease mechanisms may involve mis-regulation of the pathway genes en masse at the genetic level.

Pathway ID	Pathway Name	Ensembl[a] Gene #	Meta-brain cis eQTLs
			p<5e-03
GO:51568	**histone H3-K4 methylation**	**15**	**0.00339966**
GO:16571	**histone methylation**	**55**	**0.0019998**
P04393	Ras_Pathway	68	1
P00003	Alzheimer_disease-amyloid_secretase_pathway	66	0.89621
GO:45216	cell-cell junction organization	111	0.9976
P00052	TGF-beta_signaling_pathway	90	0.879612
GO:14069	postsynaptic density	100	0.722728
GO:34968	**histone lysine methylation**	**44**	**0.00019998**
P00010	B_cell_activation	59	0.958
GO:43414	**macromolecule methylation**	**132**	**0.00529947**

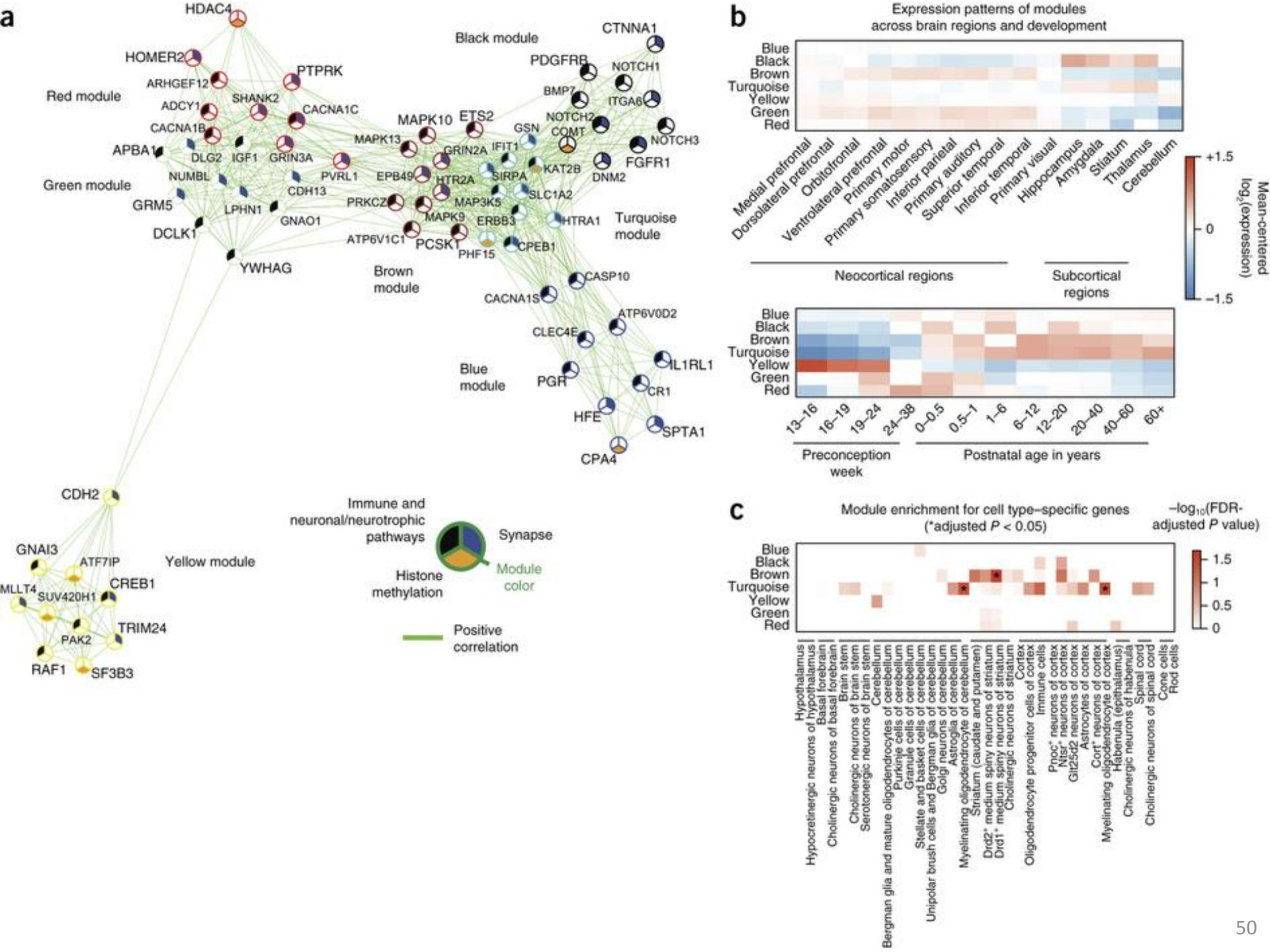
Brain Gene Co-expression Networks



- PGC pathways pipeline identified GWA enriched pathways
- Used genes in pathways to perform a supervised co-expression network analysis
- Characterized subsets of genes based on co-expression (the modules) for temporal specificity, regional specificity, cell-type specificity



Neelroop Parikshak

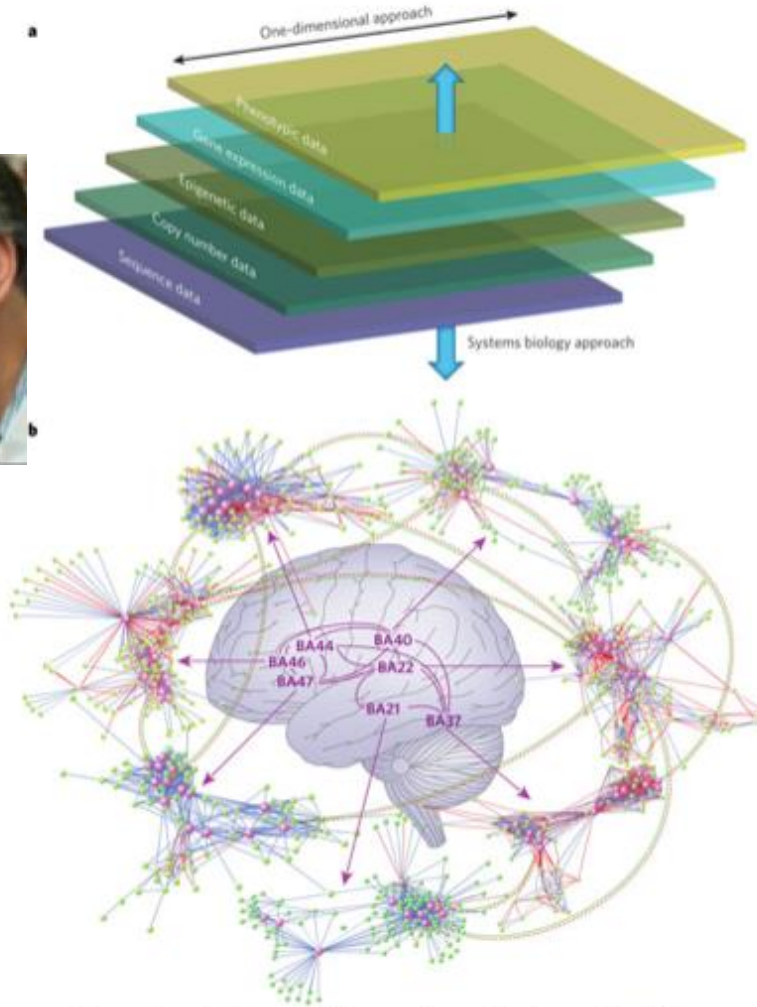

a
Red module
Green module
Black module
Blue module
Brown module
Turquoise module
Yellow module
HDAC4
HOMER2
ARHGEF12
PTPRK
ADCY1
SHANK2
CACNA1C
CACNA1B
APBA1
DLG2
IGF1
GRIN3A
PVRL1
NUMBL
CDH13
GRM5
LPHN1
GNAO1
DCLK1
YWHAG
MAPK10
ETS2
MAPK13
GRIN2A
IFIT1
EPB49
HTR2A
SIRPA
KAT2B
PRKCZ
MAP3K5
SLC1A2
MAPK9
ERBB3
HTRA1
ATP6V1C1
PCSK1
PHF15
CPEB1
CTNNA1
PDGFRB
NOTCH1
BMP7
ITGA6
GSN
NOTCH2
COMT
NOTCH3
FGFR1
DNM2
CASP10
CACNA1S
ATP6V0D2
CLEC4E
IL1RL1
PGR
CR1
HFE
SPTA1
CPA4
CDH2
GNAI3
ATF7IP
CREB1
MLLT4
SUV420H1
TRIM24
PAK2
RAF1
SF3B3
Immune and neuronal/neurotrophic pathways
Synapse
Histone methylation
Module color
Positive correlation
b
Expression patterns of modules across brain regions and development
Blue
Black
Brown
Turquoise
Yellow
Green
Red
Medial prefrontal
Dorsolateral prefrontal
Orbitofrontal
Ventrolateral prefrontal
Primary motor
Primary somatosensory
Inferior parietal
Primary auditory
Superior temporal
Inferior temporal
Primary visual
Hippocampus
Amygdala
Striatum
Thalamus
Cerebellum
+1.5
0
−1.5
Mean-centered log2(expression)
Neocortical regions
Subcortical regions
13–16
16–19
19–24
24–38
0–0.5
0.5–1
1–6
6–12
12–20
20–40
40–60
60+
Preconception week
Postnatal age in years
c
Module enrichment for cell type–specific genes (*adjusted P < 0.05)
−log10(FDR-adjusted P value)
1.5
1
0.5
0
Hypothalamus
Hypocretinergic neurons of hypothalamus
Basal forebrain
Cholinergic neurons of basal forebrain
Brain stem
Cholinergic neurons of brain stem
Serotonergic neurons of brain stem
Cerebellum
Bergmann glia and mature oligodendrocytes of cerebellum
Purkinje cells of cerebellum
Granule cells of cerebellum
Stellate and basket cells of cerebellum
Unipolar brush cells and Bergman glia of cerebellum
Golgi neurons of cerebellum
Astroglia of cerebellum
Myelinating oligodendrocyte of cerebellum
Striatum (caudate and putamen)
Drd2+ medium spiny neurons of striatum
Drd1+ medium spiny neurons of striatum
Cholinergic neurons of striatum
Cortex
Oligodendrocyte progenitor cells of cortex
Immune cells
Pnoc+ neurons of cortex
Ntsr+ neurons of cortex
Glt25d2 neurons of cortex
Astrocytes of cortex
Cort+ neurons of cortex
Myelinating oligodendrocyte of cortex
Habenula (epithalamus)
Cholinergic neurons of habenula
Spinal cord
Cholinergic neurons of spinal cord
Cone cells
Rod cells

Categories of "functional" genomics

- Differential gene expression (DGE) and co-expression gene lists related to:
  - Brain regions
  - Developmental time points
  - Cell types
  - Diseases

- Molecular regulatory relationships
  - Transcriptional and chromatin regulator binding targets
  - RNA binding protein targets

- Genes identified by other avenues
  - Genes under evolutionary constraint
  - Regions of differential methylation
  - ChIP-seq regulated regions will come soon



Geschwind and Konopka, *Nature* 2009

Assembly of Information

- DGE and co-expression gene list sources:
  - Region and time points: Kang et al., 2011* (spans 8 weeks post-conception to > 60 years of age, 16 brain regions, few individuals per time point); BrainSpan data (subset of Kang et al. samples)
  - Regions only: Hawrlycz et al., 2012 (few individuals, 100s of regions)
  - cis-eQTLs: Ramasamy et al., 2014 (10 brain regions, N ~ 100); CommonMind data (hundreds of individuals, many SCZ)
  - Cortical laminae: Miller et al., 2014 (laser capture microdissection [LCM] in fetal human); Bernard et al, 2012 (LCM in macaque)
  - Cell types: Doyle et al., 2008* (bacTRAPin mouse); Zhang et al., 2014 (FACS in mouse); Molyneux et al., 2015 (RNA-seq of fixed cells in mouse)
  - Regulators: Darnell et al., 2011 (FMRP in mouse); Weyn-Vanhentenryck et al., 2014 (RBFOX1 in mouse); Sugathan et al., 2014 (CHD8 in human)… depending on criteria, there can be many more

*Used in Nat Neuro PGC pathways paper

From human unless otherwise specified

Geschwind lab has used most of these lists in previous work, can verify they are high quality.

Disease gene lists: a far more mixed picture

Disease	Subjects		Brain Region	Platform	Study
	Cases	Ctrl			
ASD	14	14	BA9, BA41	Illumina Ref8 v3	Voineagu et al., *Nature* 2011 474(7351):380-4
	5	6	BA41/42	Affy HG-U133 plus2	Garbett et al., *Neurobiol Dis* 2008 30(3):303
SCZ	50	44	Parietal cortex	Affy HuGene 1.0 ST	Chen et al., *Mol Psychiatry* 2013 18(12):1308-14
	12	16	BA46	Affy HG-U133 plus2	Lanz (GSE53987)
	24	21	BA10	Affy HG-U133 plus2	Maycox et al., *Mol Psychiatry* 2009 14(12):1083-94
	32	29	BA46	Affy HG-U133A	Iwamoto et al., *Hum Mol Genet* 2005 14(2):241-53
	23	28	BA46	Affy HG-U133 plus2	Narayan et al., *Brain Res* 2008 1239:235-48.
BD	36	44	Parietal cortex	Affy HuGene 1.0 ST	Chen et al., *Mol Psychiatry* 2013 18(12):1308-14
	16	16	BA46	Affy HG-U133 plus2	Lanz (GSE53987)
	23	29	BA46	Affy HG-U133A	Iwamoto et al., *Hum Mol Genet* 2005 14(2):241-53
MDD	17	16	BA46	Affy HG-U133 plus2	Lanz (GSE53987)
	51	52	BA9, BA25	Affy HG-U133 plus2	Chang LC et al., *PLoS One* 2014 9(3):e90980
AD	41	36	BA3, BA11	Affy HG-U133 plus2	Blair et al., *J Clin Invest* 2013 123(10):4158-69
ETOH	17	15	Superior frontal cortex	Illumina HumanHT-12 V3	Ponomarev et al., *J Neurosci* 2012 32(5):1884

Suggested approach to integrate functional gene sets

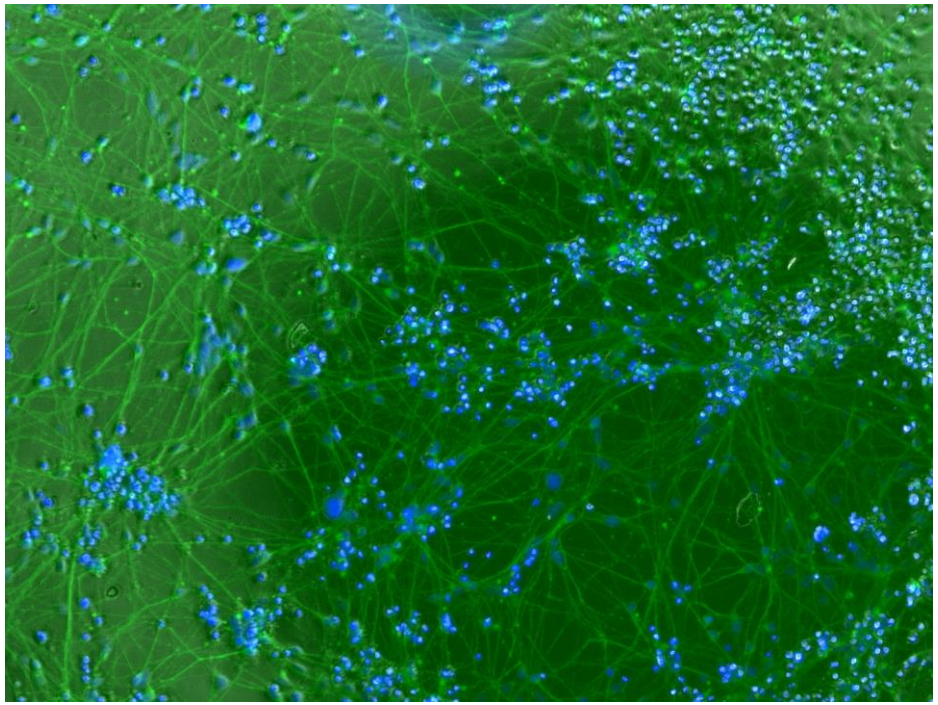
- Initial phase:
  - Use pre-computed (from Supplemental Methods of respective papers) or Geschwind lab generated lists from datasets in slide 6 as pathways in PGC pathways part 2 (code will be provided for future lists)
  - Get SCZ specific networks and lists for PGC pathways paper on SCZ?
  - Utilize Mike's cross-disorder gene expression lists and networks to perform a pathway analysis for his manuscript using the existing PGC data?
- Future suggestions:
  - Re-process large datasets from raw data for more uniform results, include newer RNA-seq datasets
  - Define modules in co-expression networks using shared genetic signals

Collaboration opportunities?

- **Improving the ontology**
  - **Adding or removing synapse-related concepts**
  - **Adding or removing relationships between concepts**
- **Improving the annotations**
  - **Adding (and even removing) annotations for genes**

Paul Thomas - GO

Connectivity Map – NIH LINCS Program



“links gene patterns associated with disease to corresponding patterns produced by drug candidates”

Neuronal Stem Cell Model Picture Opposite is from one well of a 384 cell culture plate.

Each of 5000 drugs by 3 replicates run on a cut down gene expression assay.

Data mining and comparison with disease models.

PGC Renewal

- We are writing a section called “Pathways to Therapeutics” of the PGC NIH renewal.
- Idea is work with a group of pharma in pre-competitive space on doing drug pathway analyses.
- Lilly, Pfizer, Roche, others please let me know.

Conclusions

Overall our analyses have shown the potential power of GWAS pathway analyses for the discovery of novel aetiology.

Immune and histone methylation findings show how genetic risk aggregates in pathways that may underlie vulnerability to environmental risk factors.

Using Brain Expression data may be a strong complementary approach.

Collaborations with GO and similar resources.

Acknowledgments

Thank you for listening.

The PGC, Pat Sullivan, the PGC Pathway Analysis Group, part.

Liz Rossin, Colm O`Dushlaine, Phil Lee, Laramie Duncan and Peter Holmans.

The UK NIHR and MRC for funding.