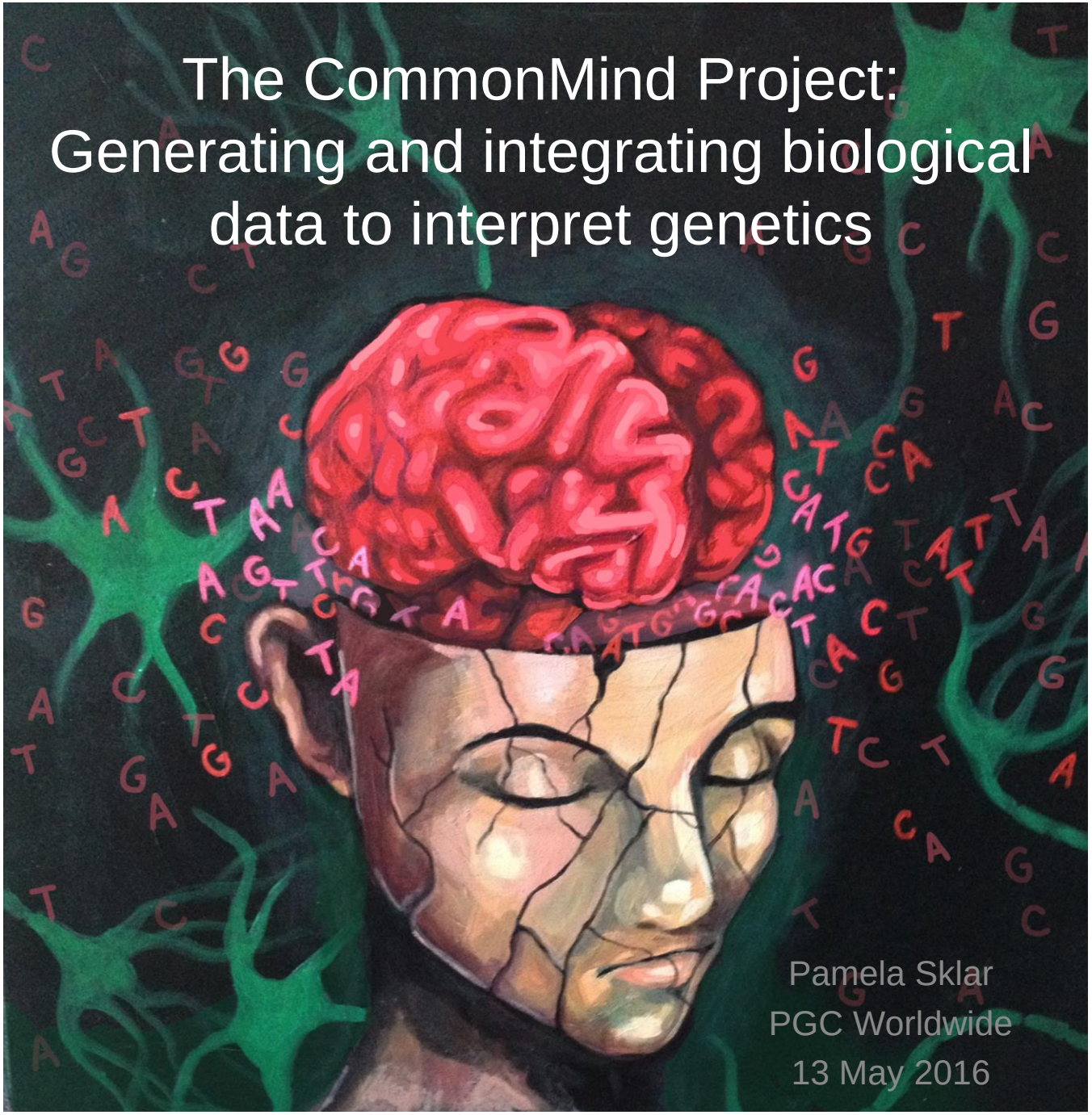
The CommonMind Project: Generating and integrating biological data to interpret genetics



Pamela Sklar
PGC Worldwide
13 May 2016

The Mystery of Schizophrenia: Inherited but Common

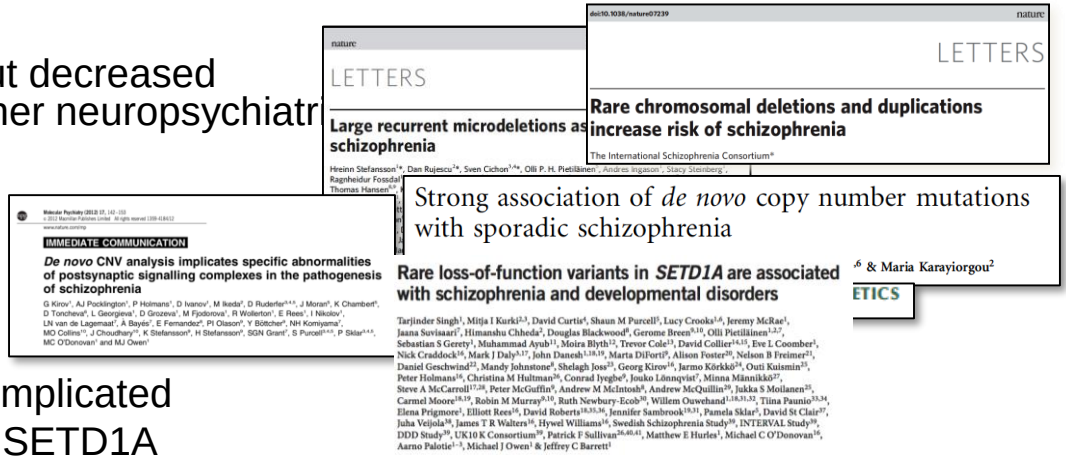
Families

Twins

Genetic risk for schizophrenia

- Common variants
  - Polygenic risk throughout the genome
  - 108 common risk loci for schizophrenia, but no single gene (mechanistically) implicated (LD)
  - 2016: C4, AS3MT

- Copy number variants
  - Increased rate in SCZ vs. controls, but decreased penetrance for CNV pleiotropic for other neuropsychiatr disease
  - Implicate synaptic genes as a class

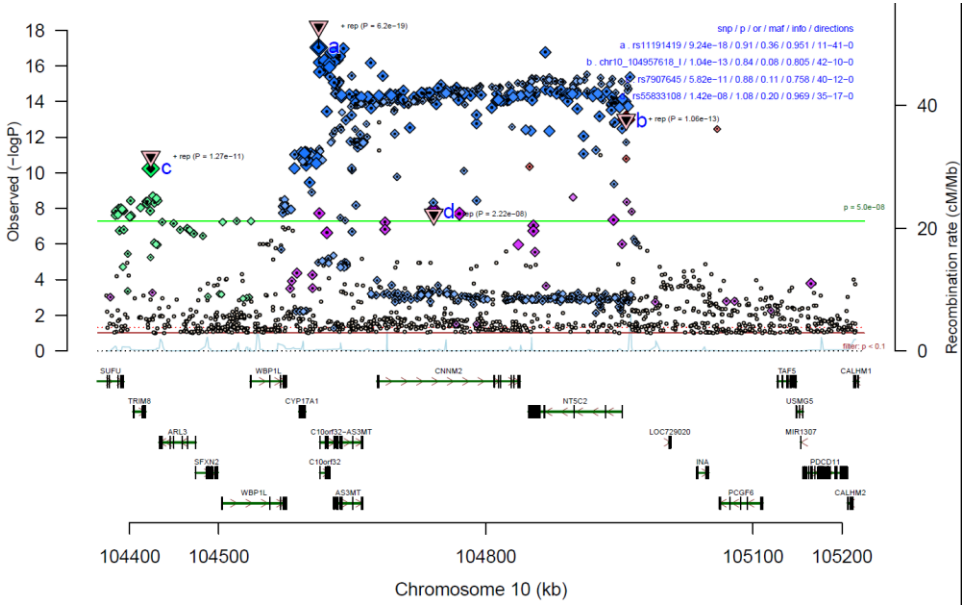


- Rare variants
  - Polygenic risk across large gene sets
  - No single variant or gene definitively implicated
  - 2016: after 20K exomes, now 1 gene SETD1A

- *De novo* variants
  - Possibly slight increase in mutation rate, but significa increase in synaptic (ARC, NMDAR) and FMRP targe genes
  - More limited role than in autism or intellectual disability, where single genes have been implicated

Which genes are candidate for follow up studies?

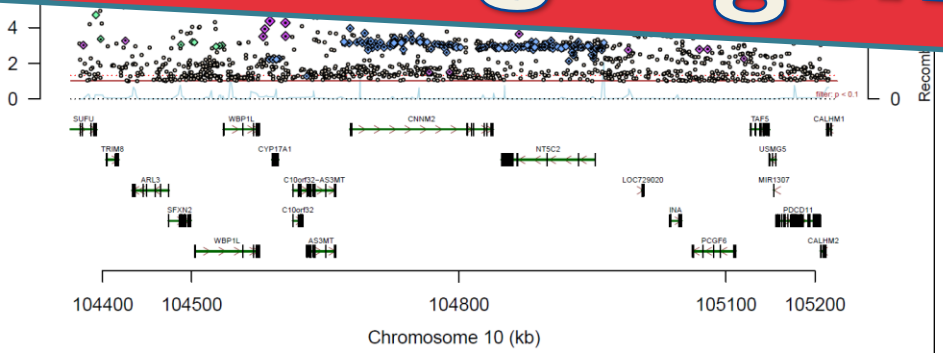
1. A proportion of SCZ risk variants lie within intergenic regions and they are far away from coding genes
2. The associated region is often large and includes multiple genes



Which genes are candidate for follow up studies?

1. A proportion of SCZ risk variants lie within intergenic

Has been difficult for genome-wide studies to nominate single genes…



CommonMind Consortium

Molecular maps of the brain

GENE EXPRESSION

- RNA sequencing - Multiple regions (DLPFC, ACC, STG), isoforms and splicing, homogenate and single cell
- Long non-coding RNA
- Long-read transcriptome sequencing
- RNA editing - microfluidic multiplex PCR and deep sequencing

GENETICS

- Common variant genotyping
- Whole genome sequencing
- Mosaicism

EPIGENOME PROFILING

PsychENCODE Project

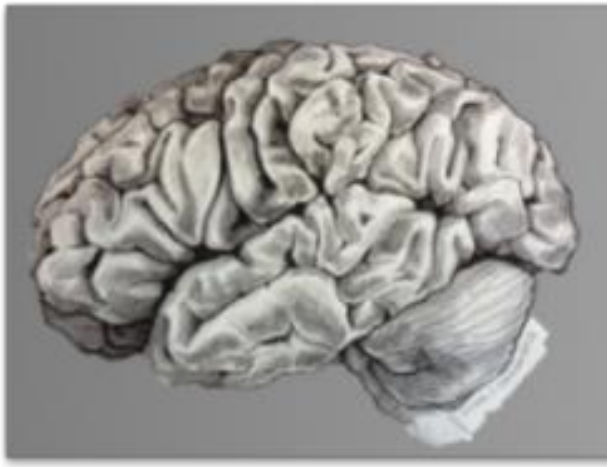
- Histone modification - Chromatin Immunoprecitation (ChIP-seq): H3K4me3, H3K27ac and others, pools of neuronal subtypes
- Chromosome conformation capture (3C) genome-wide
- DNAse sequencing to identify open chromatin - Assay for transposase accessible chromatin (ATACseq), Self-transcribing active regulatory region sequencing (STARR-seq)

PROTEOMICS

- PSD focused Liquid chromatography (LC)-selective reaction monitoring (SRM)-mass spectrometry

CommonMind Consortium Members

- **Icahn School of Medicine at Mount Sinai**
  - Pamela Sklar, Joseph Buxbaum, Schahram Akbarian, Andrew Browne, Rai Chang, Alex Charney, Genomics Core (Yumi Kasai, Violeta Capric), Menachem Fromer, Jessica Johnson, Davy Kavanagh, Harry Haroutunian, Milind Mahajan, Dalila Pinto, Shaun Purcell, Panos Roussos, Doug Ruderfer, Hardik Shah, Bin Zhang, Qingrun Zhang, Eric Schadt, Jun Zhu
- **University of Pittsburg**
  - David Lewis, Bernie Devlin, Cong Lu, Bert Klei
- **University of Pennsylvania**
  - Raquel Gur, Chang-Gyu Hahn
- **Duke & University of North Carolina**
  - Gregory Crawford, Patrick Sullivan
- **F. Hoffman-La Roche Ltd**
  - Enrico Domenici, Laurent Essioux
- **Takeda Pharmaceuticals Company Limited**
  - Keisuke Hirai



- **Sage Bionetworks**
  - Lara Mangravite, Mette Peters, Solly Sieberts, Kristen Dang, Thanneer Perumal, Ben Logsdon, Brian Bot, Mike Kellen
- **NIMH**
  - Thomas Lehner, Robin Kramer, Barbara Lipska

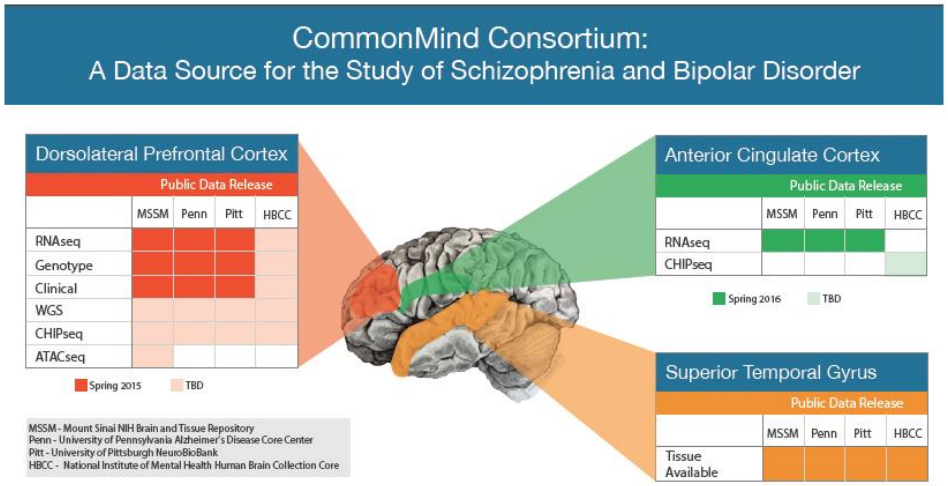
Brain Banks and Sample Selection

- Schizophrenia, schizoaffective disorder, bipolar disorder
- Generally older population, case and control ages comparable
- Exclusions: ventilator near death, AD/PD by pathology
- Total phase 1
  - N = 717
- Replication and Extension
  - Microarray data from dbGAP 000979.v1.p1

Cohort	Group	N	Age (years)
Mount Sinai (MSSM)	Schizophrenia	195	73.4
	Bipolar	24	62
	Control	204	73.5
University of Pennsylvania	Schizophrenia	62	79.2
	Control	40	67.5
University of Pittsburgh	Schizophrenia	61	47.7
	Schizophrenia Control	61	48.9
	Bipolar	35	45.5
	Bipolar control	35	46.4
NIMH Human Brain Collection Core	Schizophrenia	128	51.2
	Bipolar	89	42.5
	Control	251	34.6

Haroutounian, Lewis, Gur, Hahn, Lipska

Where to get the data?

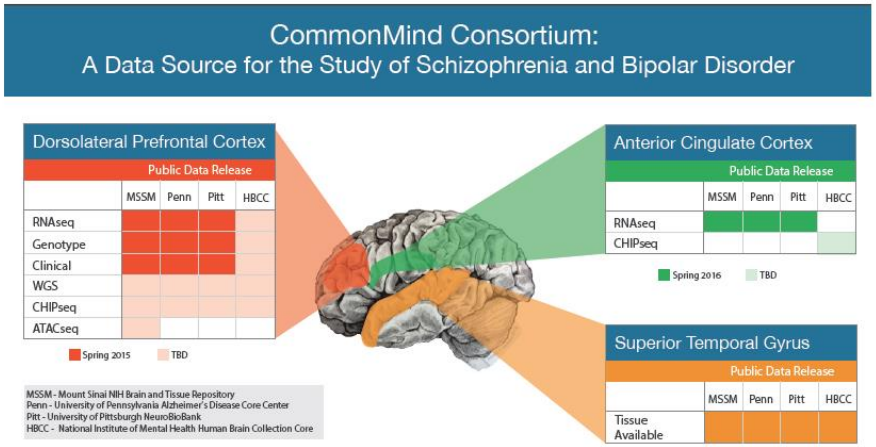


Data Release: Sage Bionetworks Synapse Platform

COMMONMIND CONSORTIUM KNOWLEDGE PORTAL

- Committed to rapid public, pre-publication release of raw data, followed by analyses, code etc
- *Release 1:* RNAseq from > 600 samples DLPFC and covariates
- *Release 1.1:* BAM files with alligned and unmapped reads, QC'd and imputed genotype data
- *January 2016:* PsychENCODE NeuN+, NeuN-, DLPFC, ACC ChiPSeq
- *Summer 2016:* methods, source code, clinical data for release 1
- *Summer 2016:* Anterior cingulate RNAseq

https://www.synapse.org/#!Synapse:syn2759792/wiki/69613

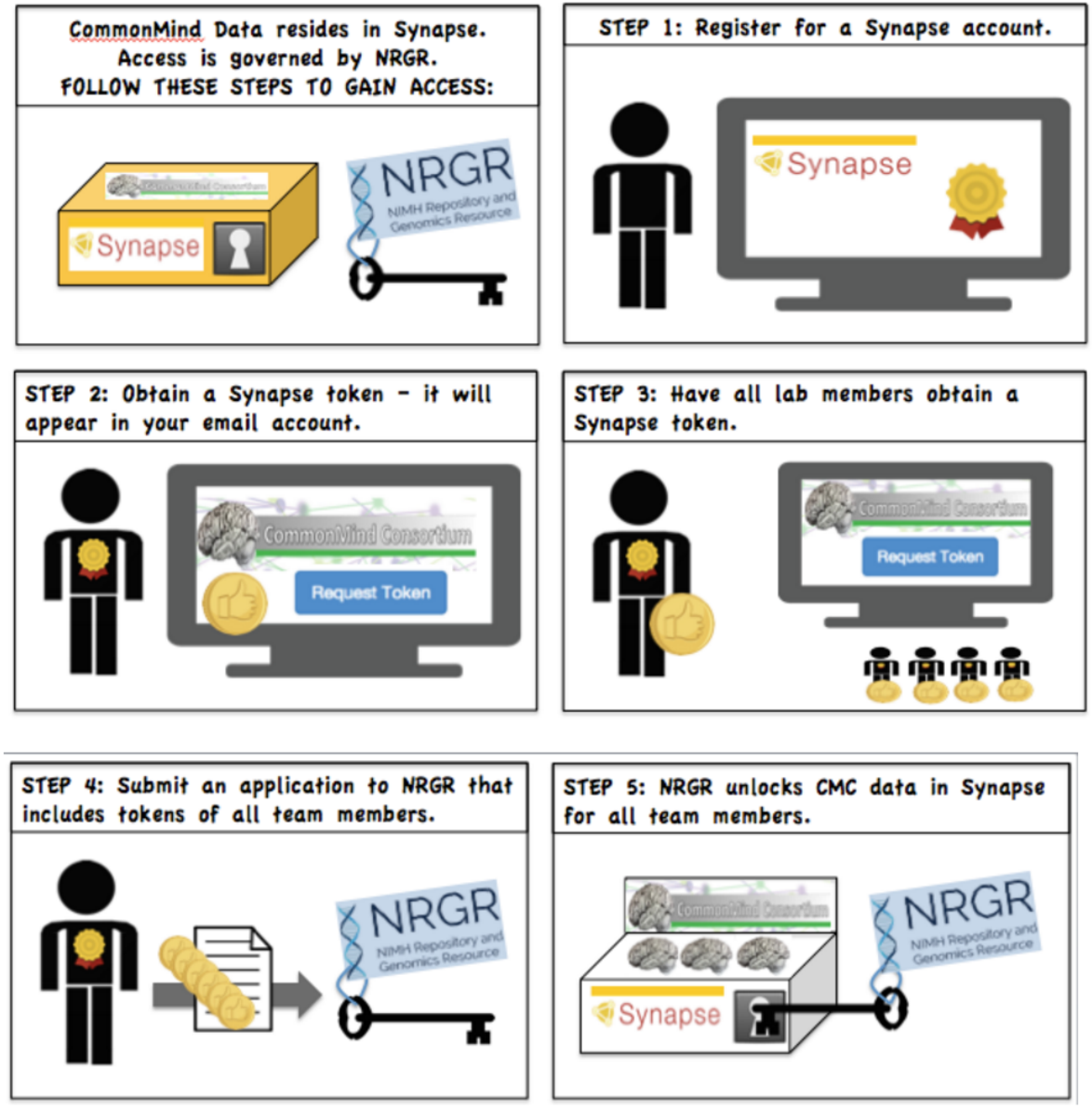


Data Release: Sage Bionetworks

COMMONMIND CONSORTIUM KNOWLEDGE PORTAL

How to find and download a file

1. **Browse content:** Content of this portal may be browsed without restrictions and browsing files does not require a Synapse account.
2. **Register for a Synapse Account:** In order to get access to the data you must first register with Synapse.
3. **Fulfill Data Use Terms:** Access to Controlled data requires approval by the NIMH Repository and Genomics Resources (NRGR).
4. **Access Data:** After approval by NRGR you will have permission to access all Controlled data. Before downloading it you must agree to acknowledge the consortium in publications and to an embargo on the bipolar data (in effect until August 1, 2015). You must use this web client to agree to these data use terms, even if you plan to download data programmatically.
5. **Locate Files for Download**



From our FAQs:

Why do I need approval from the NIMH Repository and Genomics Resources (NRGR) to access Controlled CommonMind data?

Access to CommonMind data is monitored and controlled by the NRGR in compliance with NIH policy for protection of human research data.

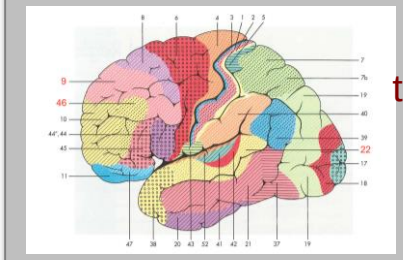
AVOIDS dbGAP access requests

Data will be available in dbGAP soon

RNA-seq experiments + Genetics QC

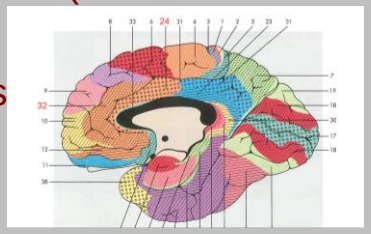
doi: http://dx.doi.org/10.1101/052209

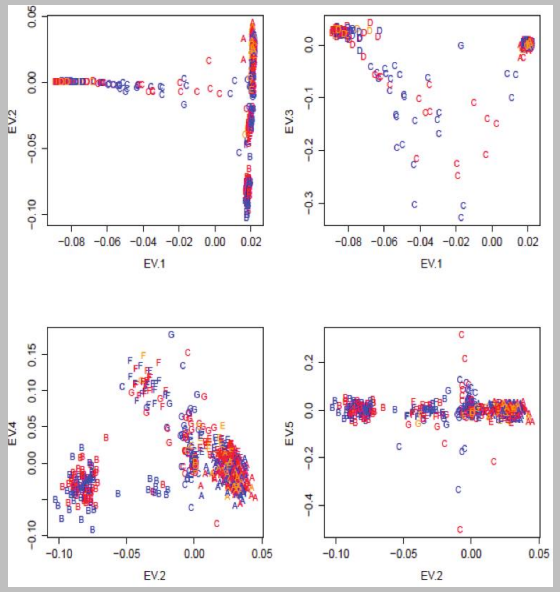
Dorsolateral prefrontal cortex (Brodmann 9 & 46)



Superior temporal gyrus (Brodmann 22)

Anterior cingulate cortex (Brodmann 24 & 32)





RED=CONTROL
BLUE=SCZ
ORANGE=BD

MSSM Brain Bank | Pitt Brain Bank | Penn Brain Bank

Samples processed in randomized order without respect to brain bank, diagnosis or other clinic al characteristcs

Single Site Processing
DNA and RNA isolation

Re-randomization in batches of 120-226

Genotyping

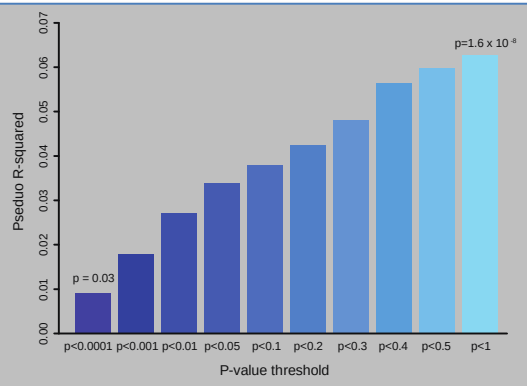
Imputation
Ancestry Evaluation
Polygenic Scoring
DNA based sample QC

Ribozero and library preparation

RNA sequencing

Outlier identification

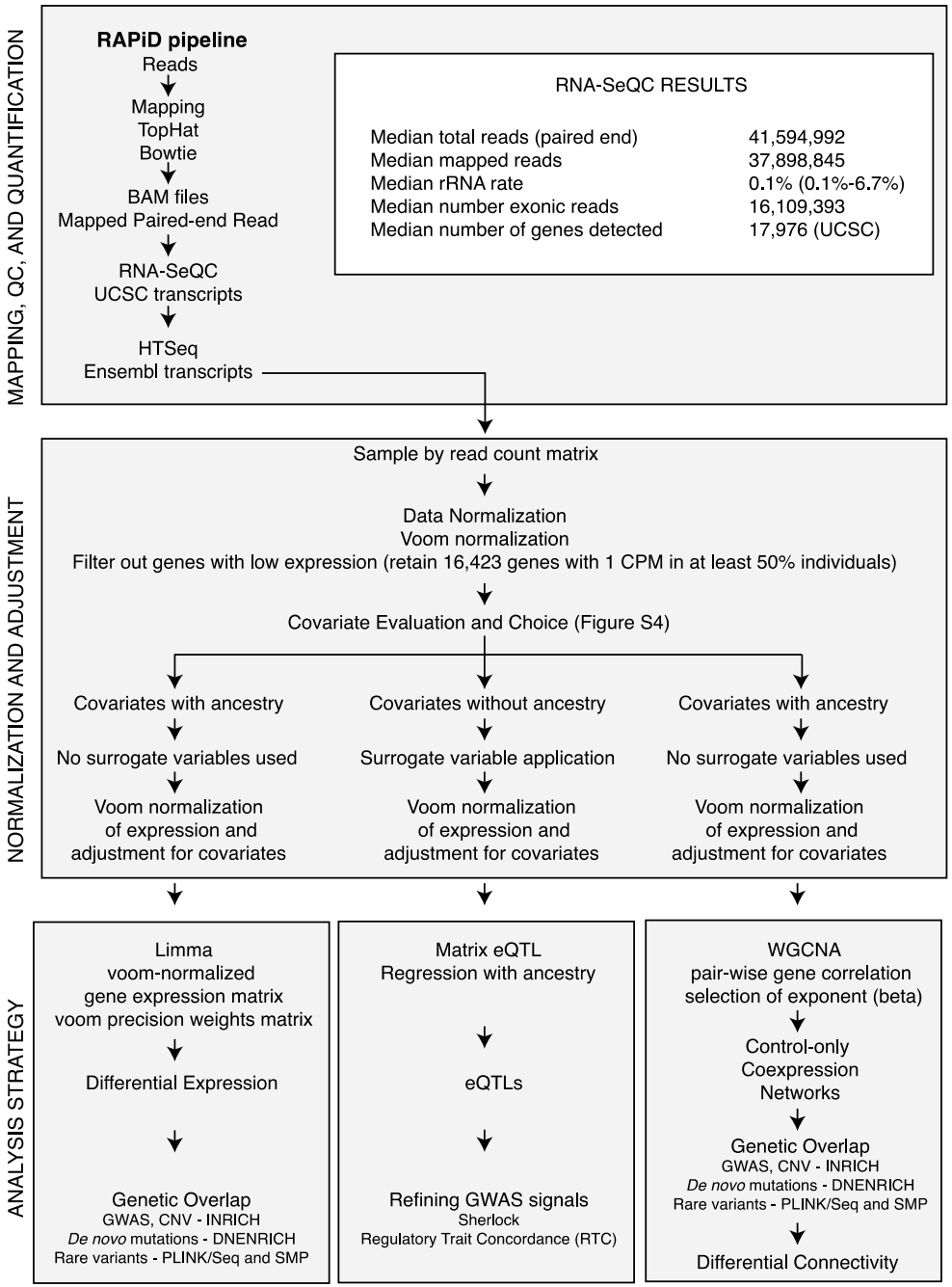
DNA/RNA matching



MAPPING, QC, AND QUANTIFICATION

NORMALIZATION AND ADJUSTMENT

ANALYSIS STRATEGY



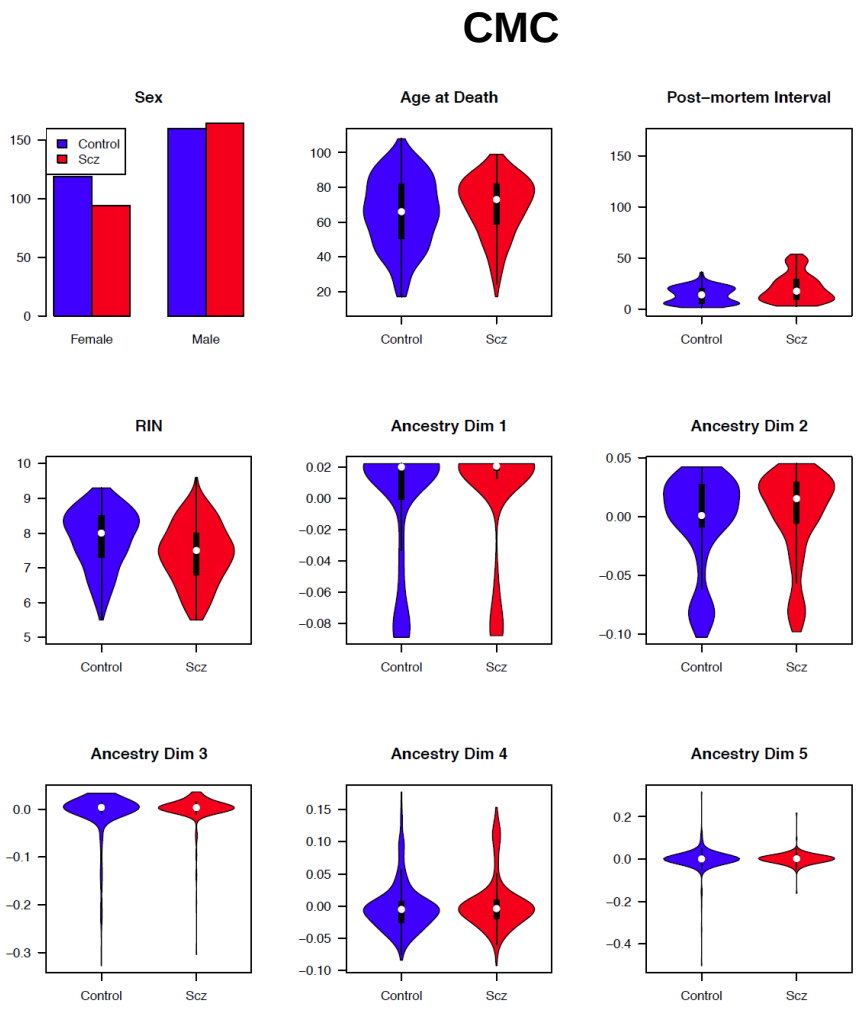
Variables in DE model

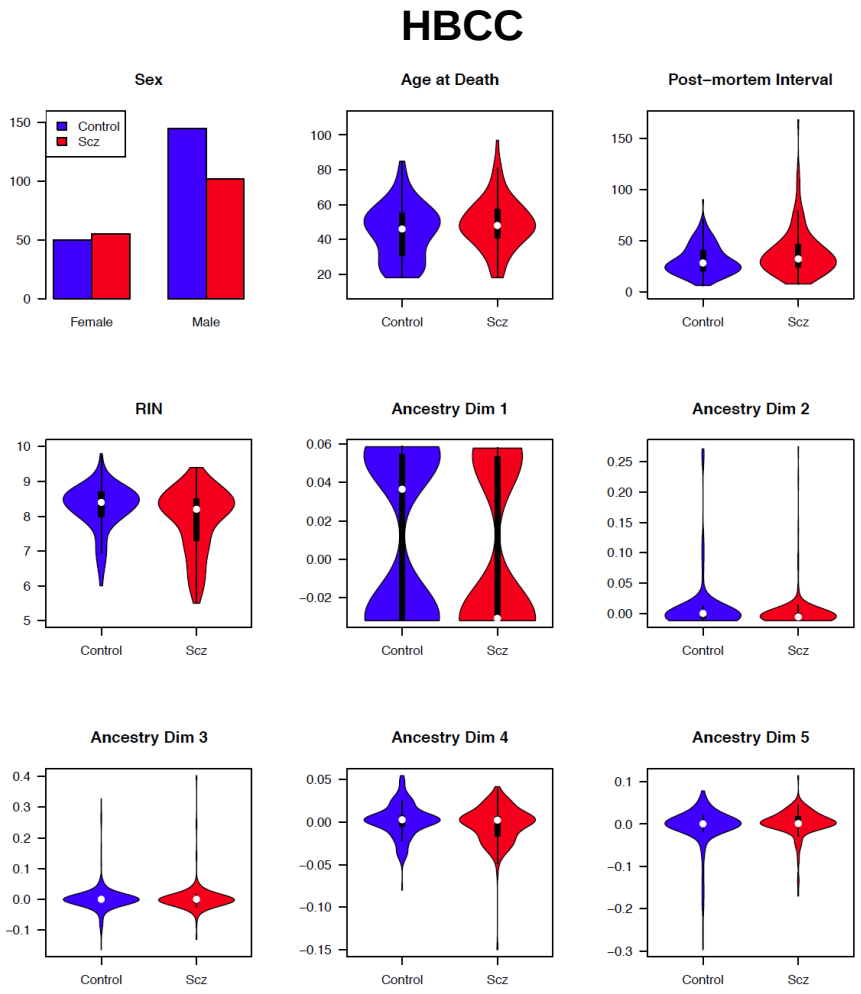
fraction of variance explained by each covariate

Effect	Median
RIN+RIN^2	0.1030
Library Batch	0.0660
Institution	0.0293
Diagnosis	0.0281
Age	0.0070
Ancestry	0.0059
Postmortem Interval	0.0026
Sex	0.0009
All known covariates	0.4206

Bert Klei
Bernie Devlin
Menachem Fromer

Distribution of Biological Covariates

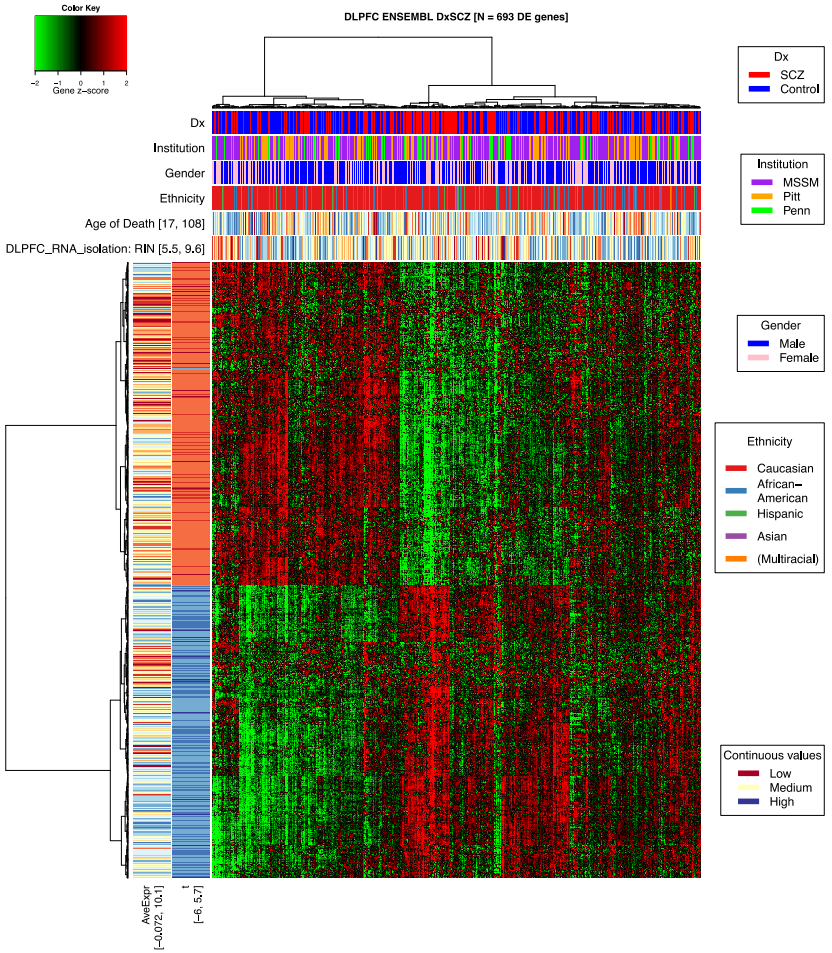




Hierarchical clustering of the final cohort

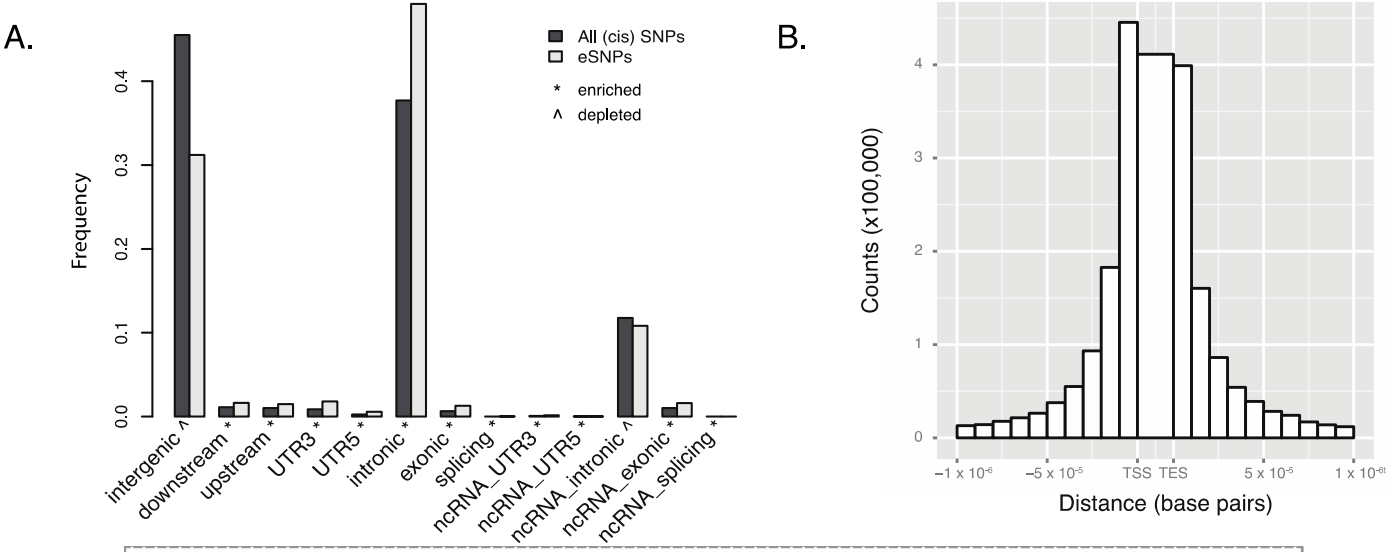
FINAL COHORT

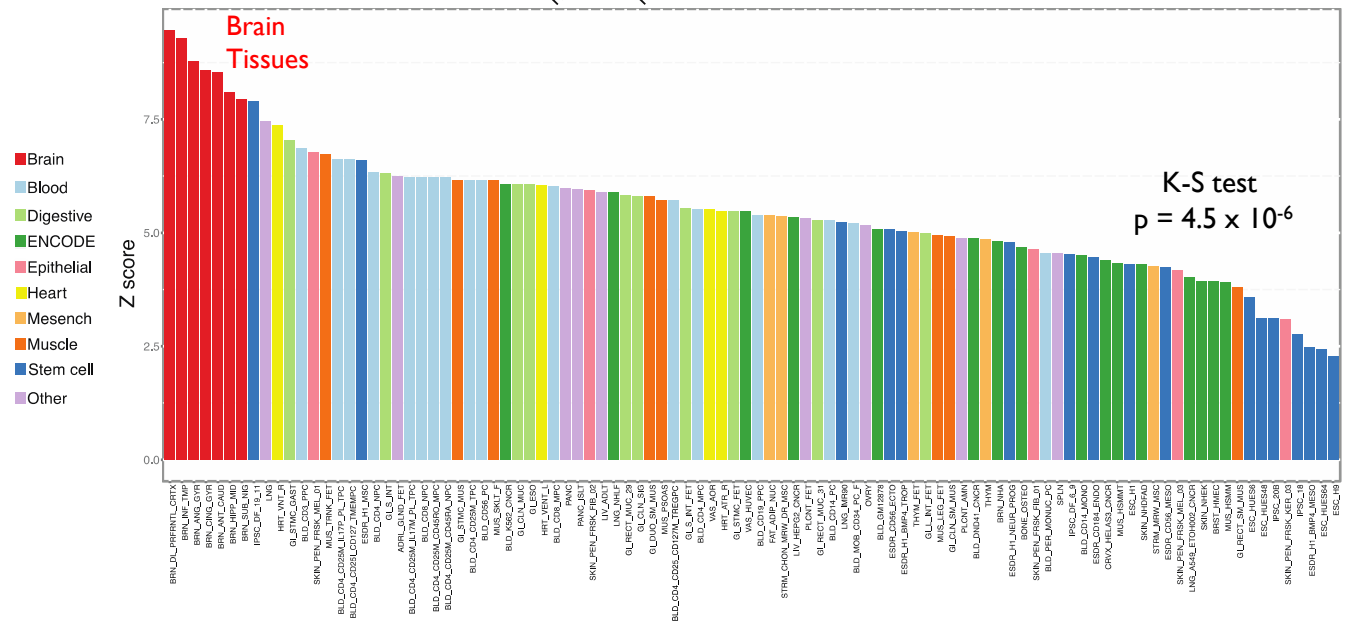
Metric	Value
Schizophrenia	258
Bipolar Disorder	55
Controls	279
Male, Female	351, 241
Caucasian	478
African-American	85
Hispanic	24
Asian	4
Median Age	67.5
Median PMI	16
Median RIN	7.7



Menachem Fromer

CMC: Largest brain eQTL resource





Source	π_1	eGENES in CMC
GTex	0.95	11,853
Harvard Brain Bank	0.75	7,275
Brain Cloud	0.70	7,180
NIH	0.79	9,995
UKBEC	0.93	11,300
Meta-analysis	0.62	9339

Solveig Sieberts
Panos Roussos

Interpreting genetics: CMC

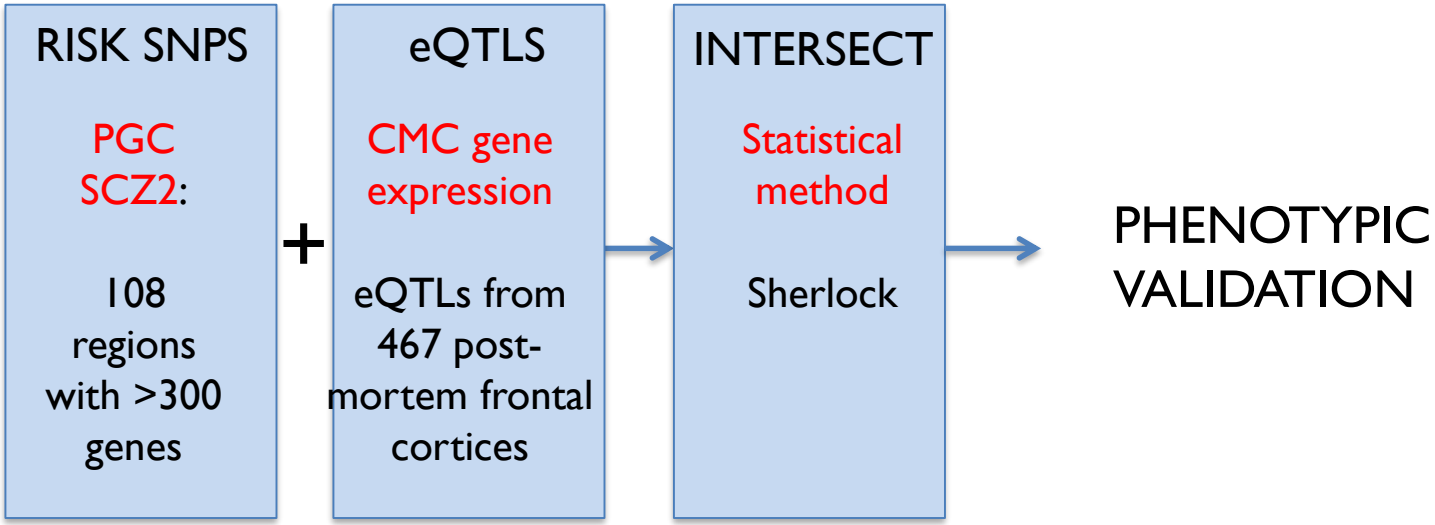
Observation:

Genetic variants alter SCZ risk (SNPs)

Genetic variants control gene expression (eQTLs)

Hypothesis:

If a risk variant also causes gene expression changes then we have identified a potentially causal biological mechanism



Edwin Oh
Mahsa Parvisi
Patrick Sullivan
Nico Katsanis

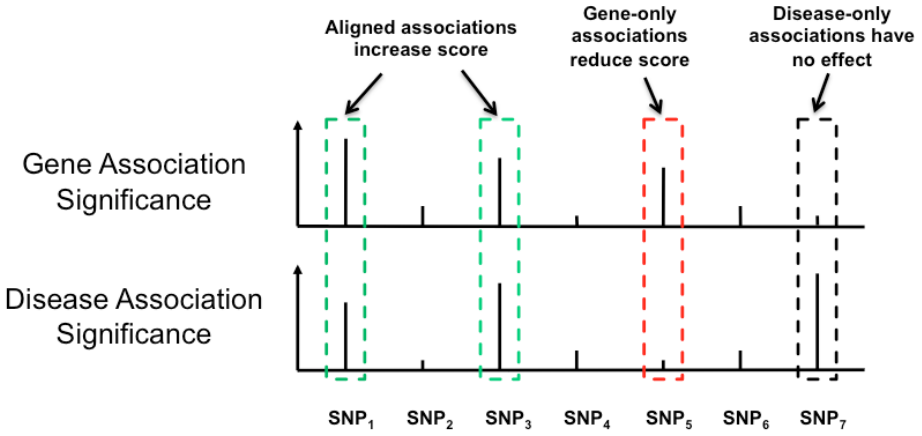
Can we identify for a single time point and tissue, instance where the disease associated GWAS SNP is convincingly associated with an expression change

Method: Sherlock

Independent eQTL (LD block, all eSNPs associated with a gene within <500kb)

Single Bayes factor calculated per block, likelihood of observed GWAS and eQTL p values under model where they mediate risk relative to the null

Combined into a per gene score, p values estimated by permutation



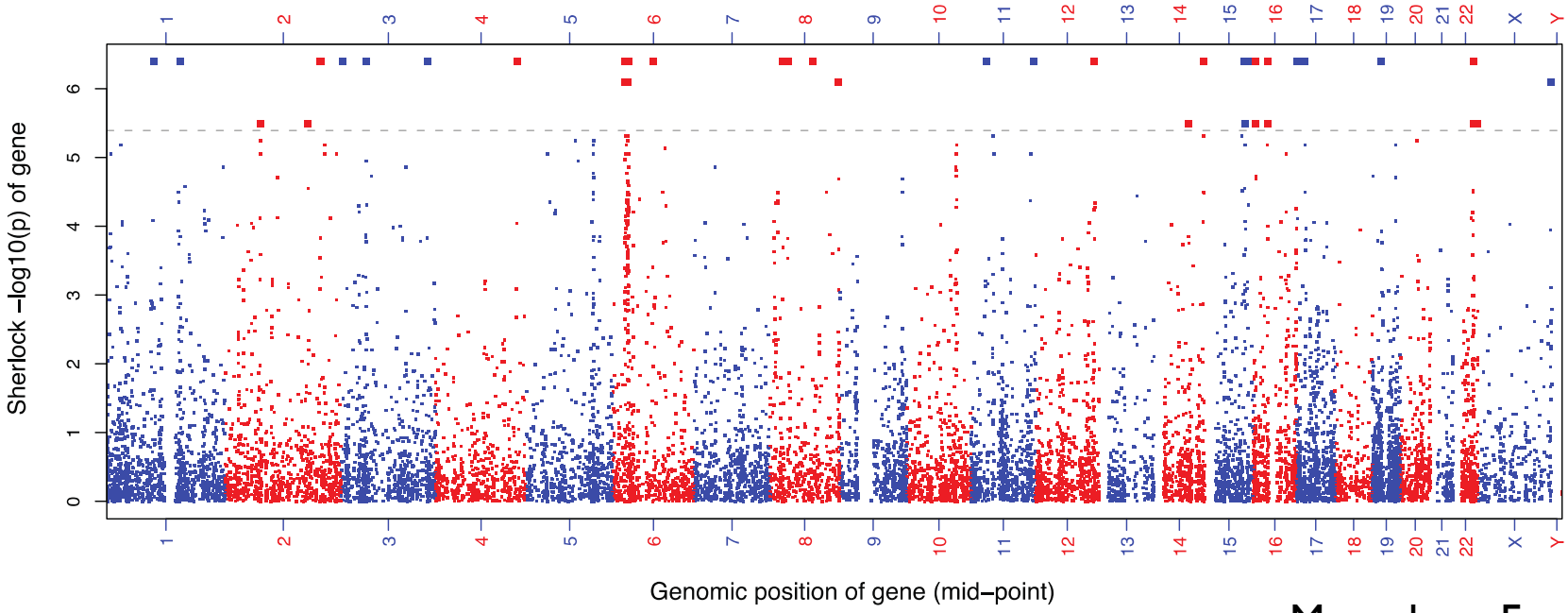
He, AJHG, 2013

Manhattan Plot

12,367 Ensembl genes with one or more eQTL with PGC SCZ2 GWAS data

Bonferroni corrected for all genes

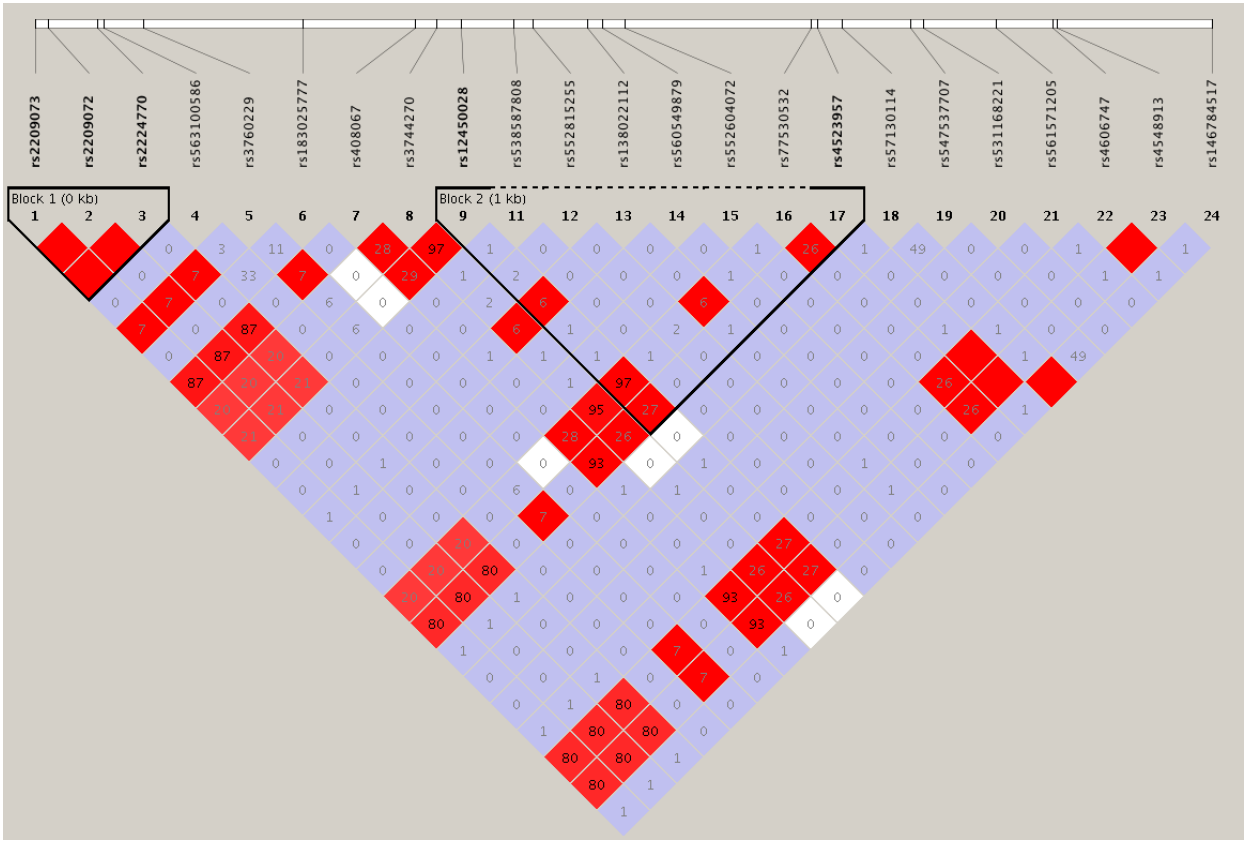
84 genes in 30 regions passed the statistical threshold



Menachem Fromer

Sherlock alone provides insufficient evidence for some loci

SRR, serine racemase



	eQTL	GWAS
rs12450028		
rs4523957		

TT .328
CG .359
CT .313

T = SCZ risk

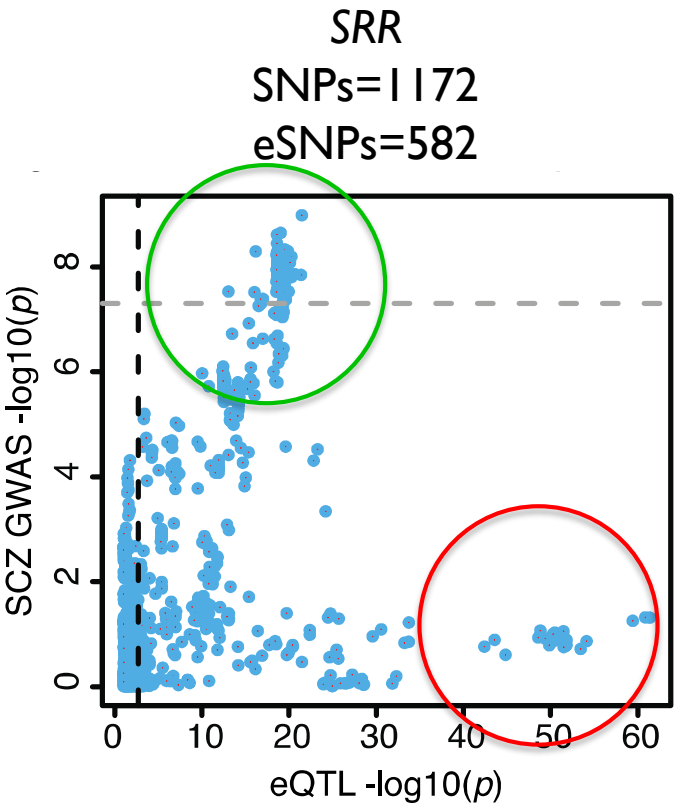
http://browser.1000genomes.org/Homo_sapiens/UserData/Haploview

SRR haplotypes and associated traits

eQTL	GWAS		
rs12450028	rs4523957		
T	T	.328	
C	G	.359	
C	T	.313	

Direction of effect for haplotype associations	
SCZ risk	SRR expression
+	-
-	+
+	+

Correlation between GWAS signals and eQTL signals



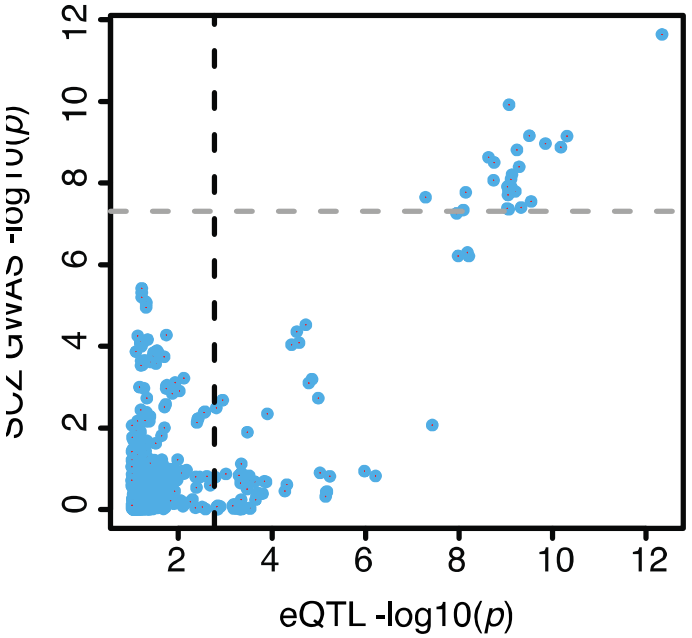
GWAS SNPs are associated with expression changes

But, best eQTLs are not associated with SCZ

Thus, disease association in a region can "predispose" to having decreased expression, even it not causally related to the disease

Additional post-Sherlock filtering

FURIN
SNPs=575
eSNPs=98



Filtered out:

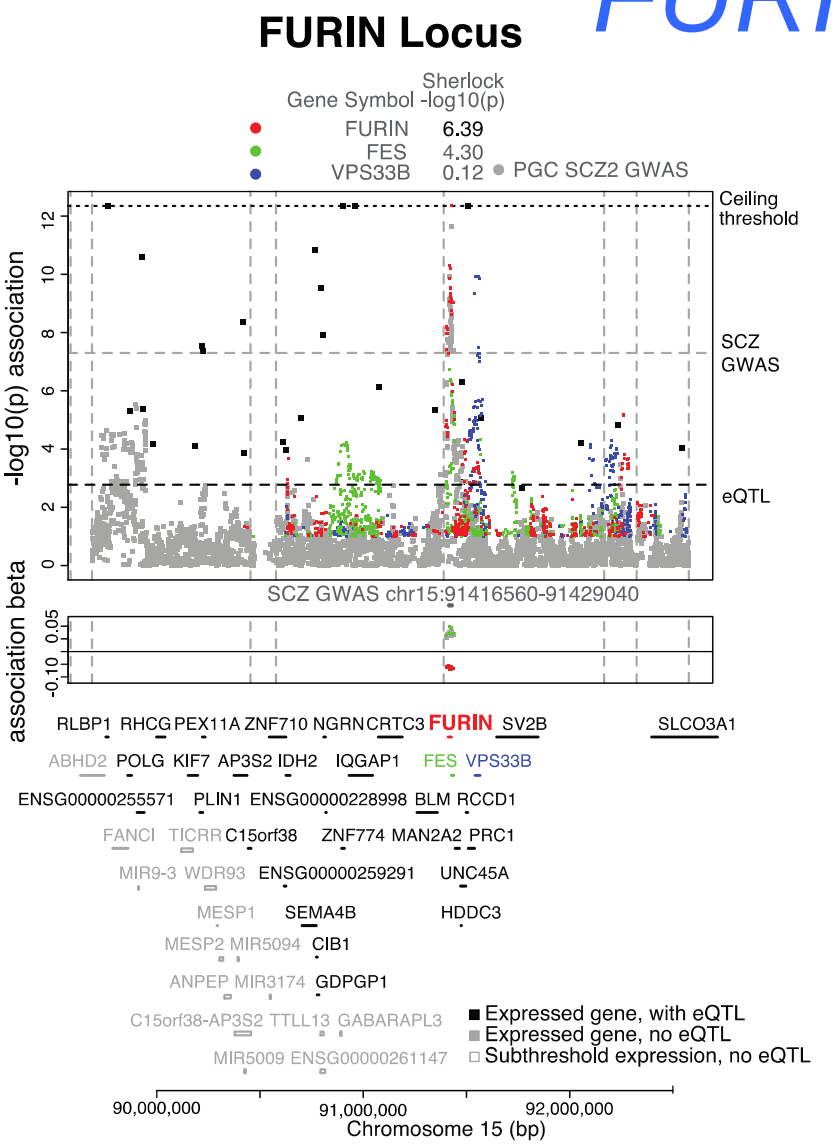
- Genes with poor correlation
- Genes where second gene in locus had a significant isoform
- Genes where the locus had additional subthreshold genes
- Only trans eQTL support

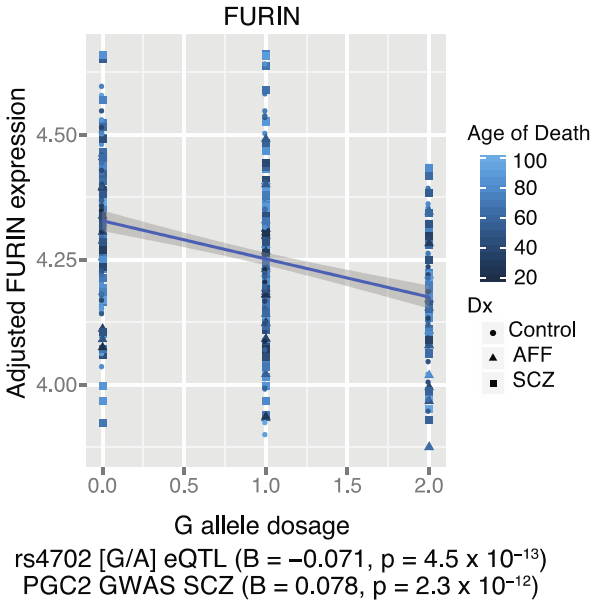
Results:

5 loci with single gene

FURIN; proprotein convertase family protein
SNAP91; synaptosomal-associated protein of 91 kDa
CLCN3; voltage-gated chloride channel 3
TSNARE1; t-SNARE domain containing 1
CNTN4; contactin 4

eQTL association profile
FURIN

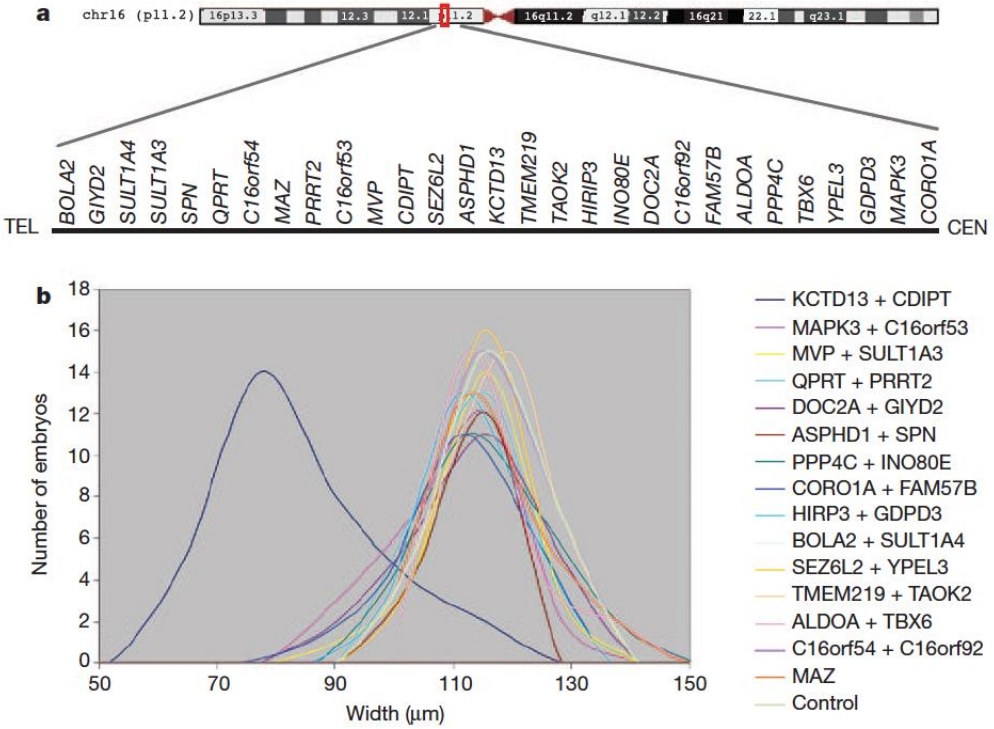


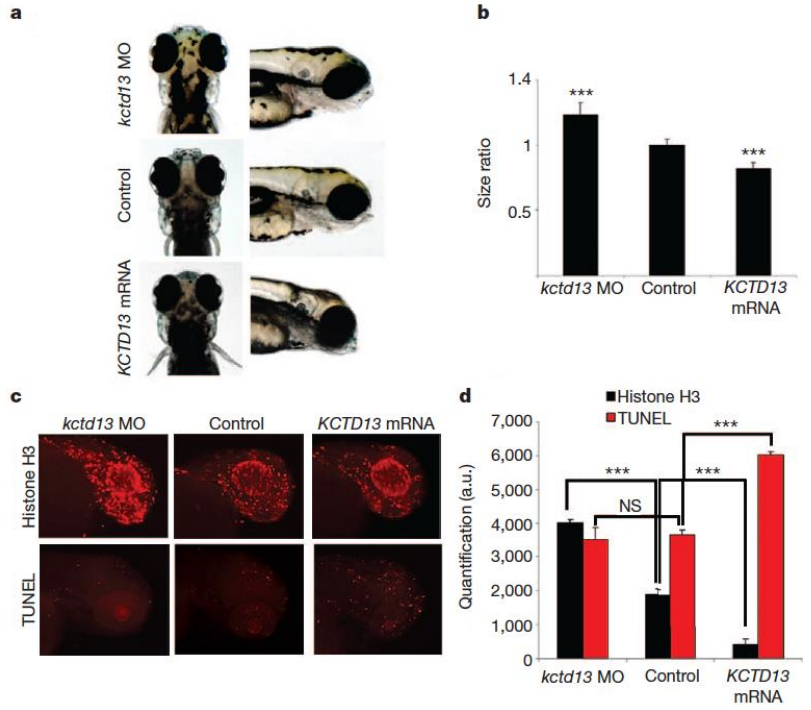


rs4702 [G/A] eQTL (B = –0.071, p = 4.5 x 10^{-13})
PGC2 GWAS SCZ (B = 0.078, p = 2.3 x 10^{-12})

Functional?
Relevant?

Modeling neuropsychiatric disorders -16p11.2 exemplar





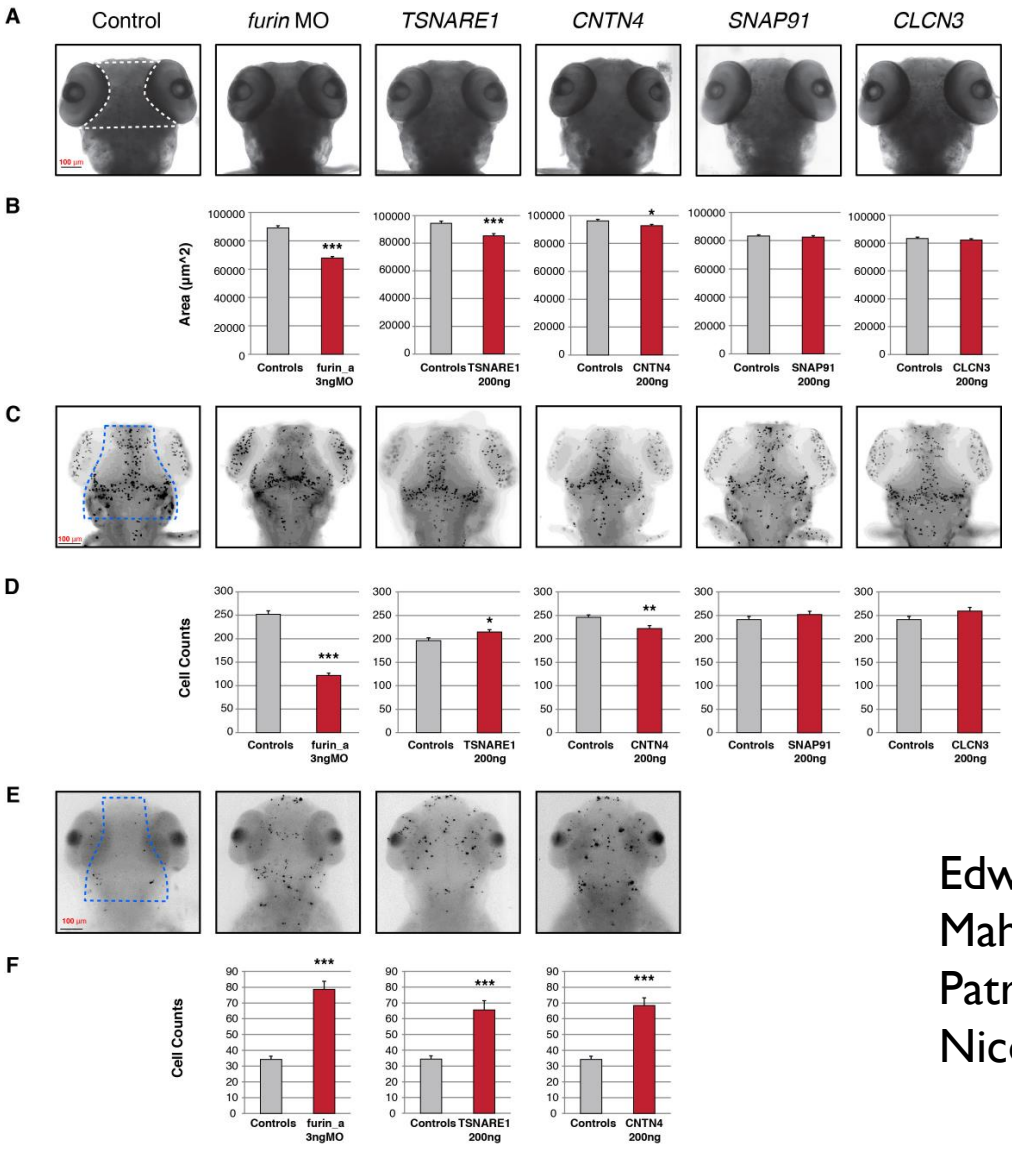
Golzio et al. Nature 2012

Modeling neuroanatomical changes

Underexpression and overexpression led to decreased head size

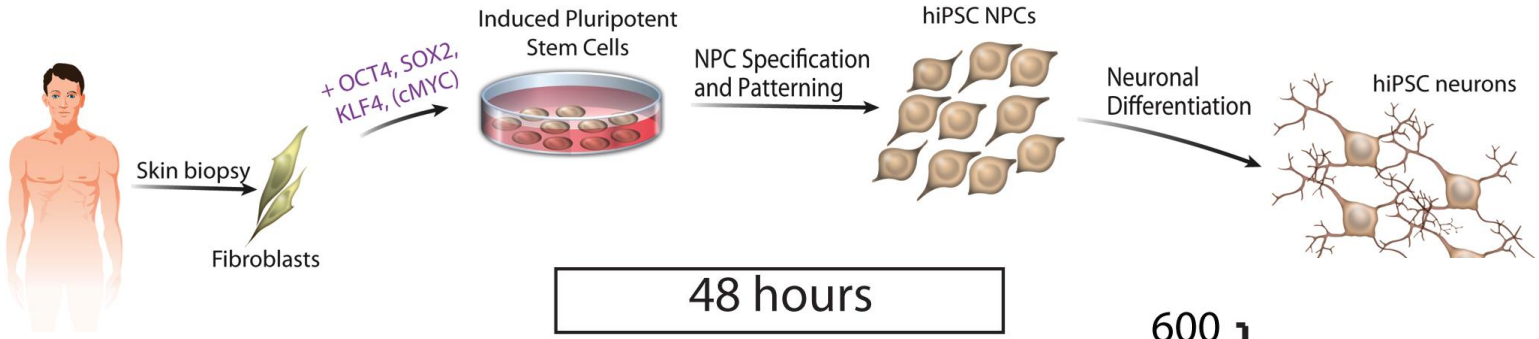
Phosphohistone 3 staining primarily shows decreased proliferation

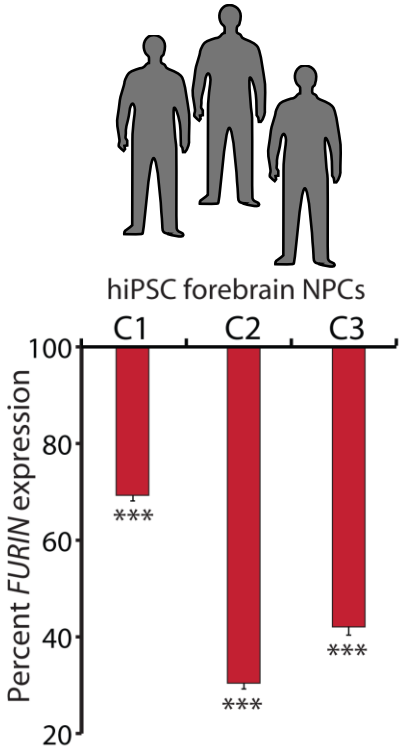
TUNEL staining shows higher apoptotic index

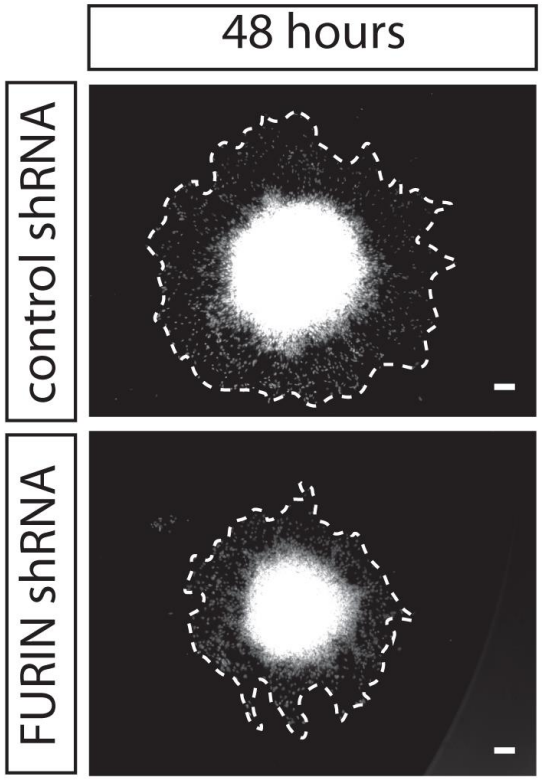


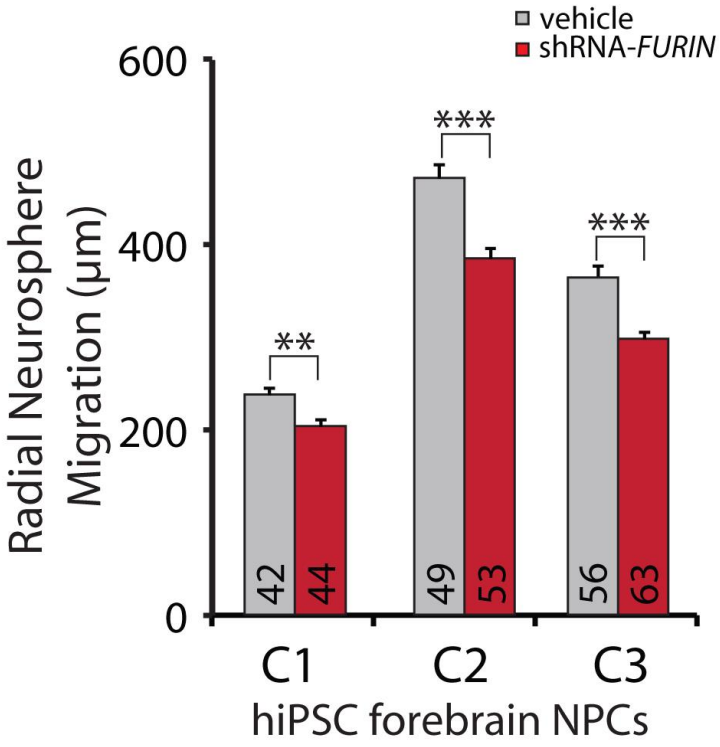
Edwin Oh
Mahsa Parvisi
Patrick Sullivan
Nico Katsanis

Decreased *FURIN* expression decreases neural migration in forebrain NPCs









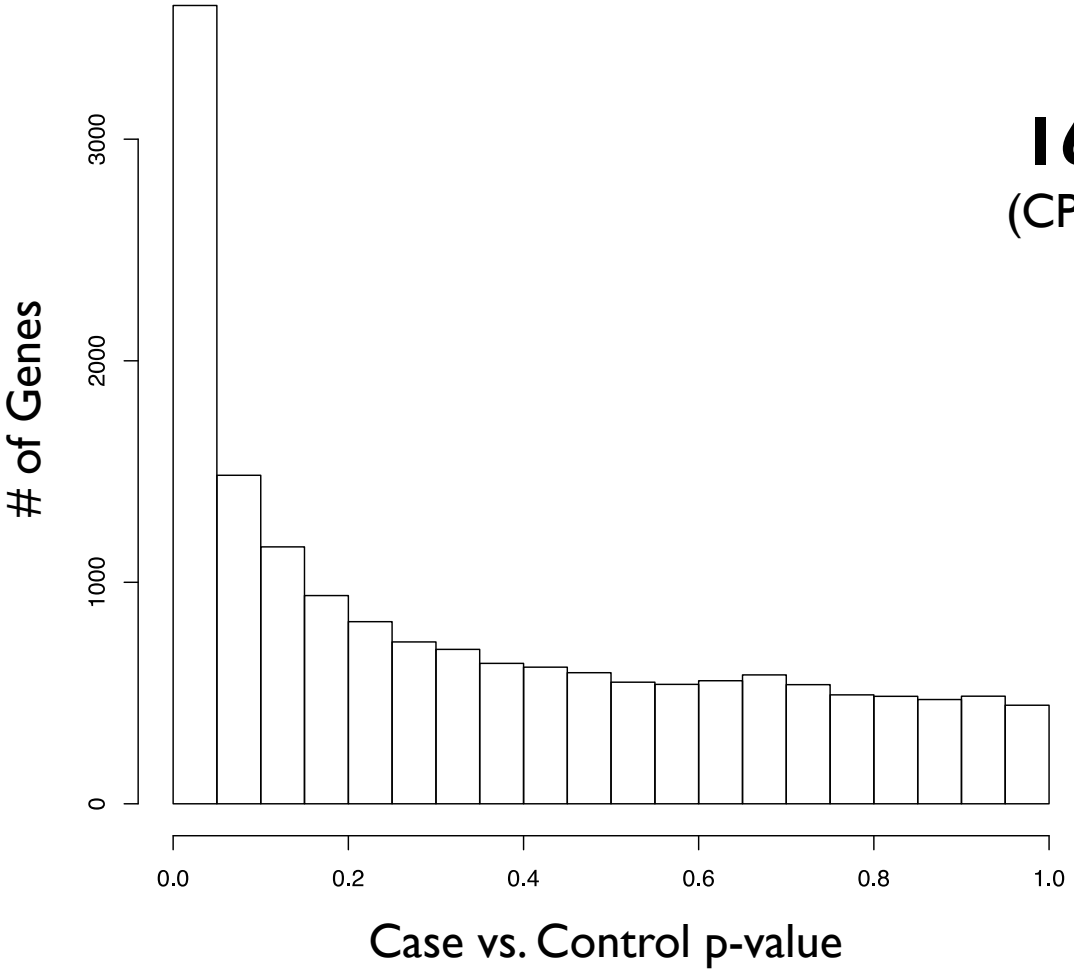
Aaron Topol

Kristen Brennand

Differential expression

Many genes differentially expressed in DLPFC…

π_1 estimate: **44% of genes are differentially expressed**



Analyzed:

16,423 Ensembl genes

(CPM > 1 in at > 50% of 592 samples)

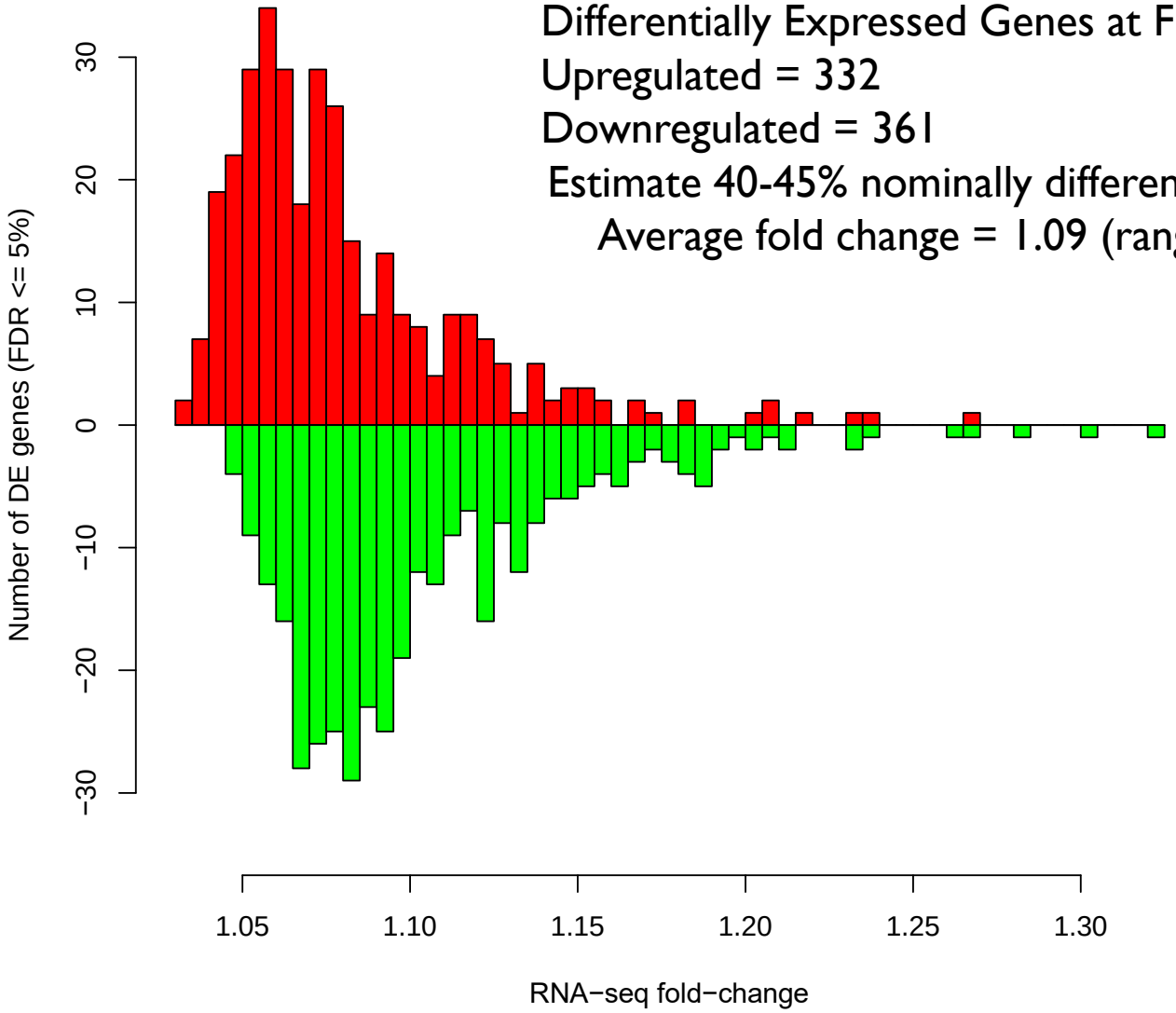
X

258 SCZ vs. 279 controls

Menachem Fromer

Storey and Tibshirani, PNAS, 2003
Mostafavi, Mol Psych, 2014

But with small effect sizes



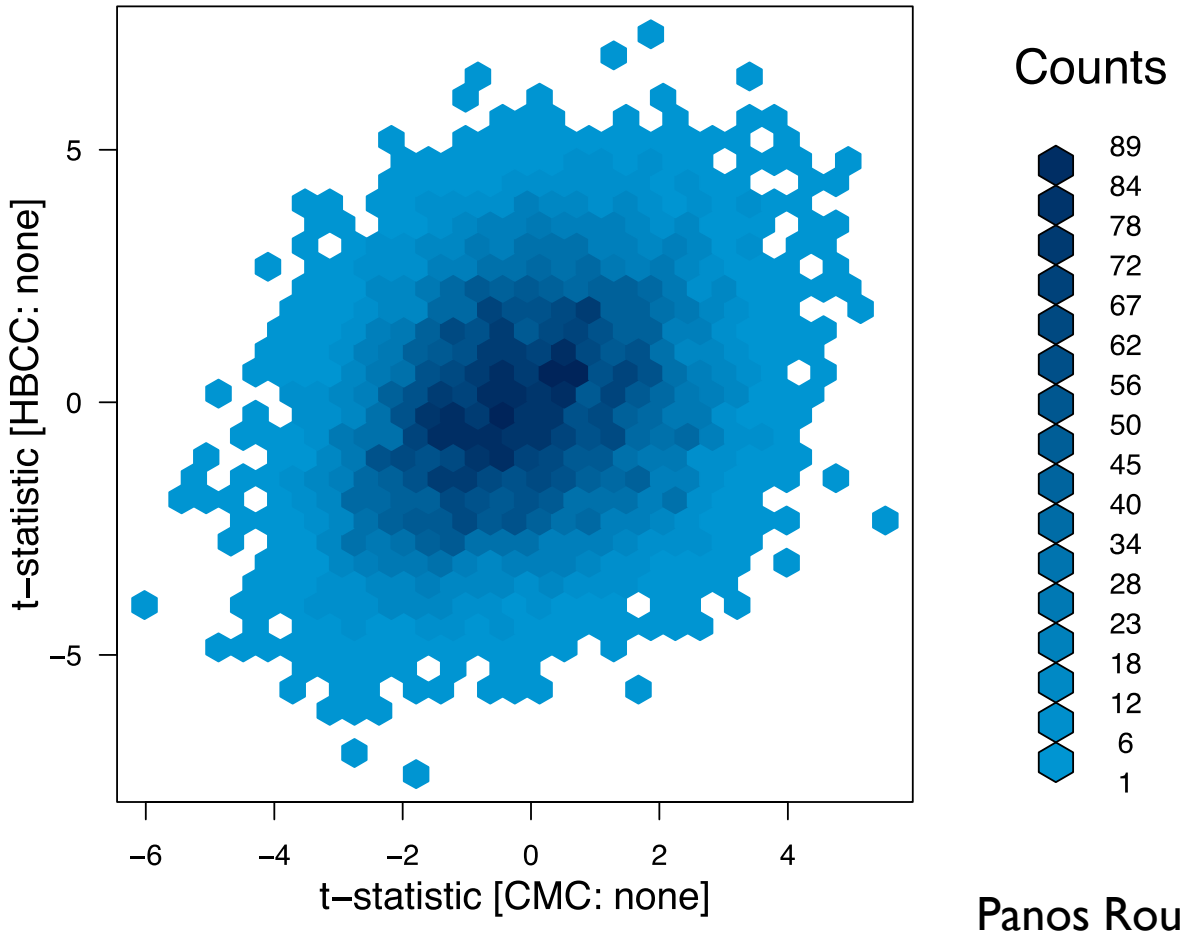
Differentially Expressed Genes at FDR $\leq$ 5% = 693
Upregulated = 332
Downregulated = 361
Estimate 40-45% nominally differentially expressed
Average fold change = 1.09 (range 1.03-1.33)

Independent samples

NIMH HBCC samples: IlluminaHT-12_V4 Beadchips

Pearson correlation
R = 0.58

131 SCZ and 176 control

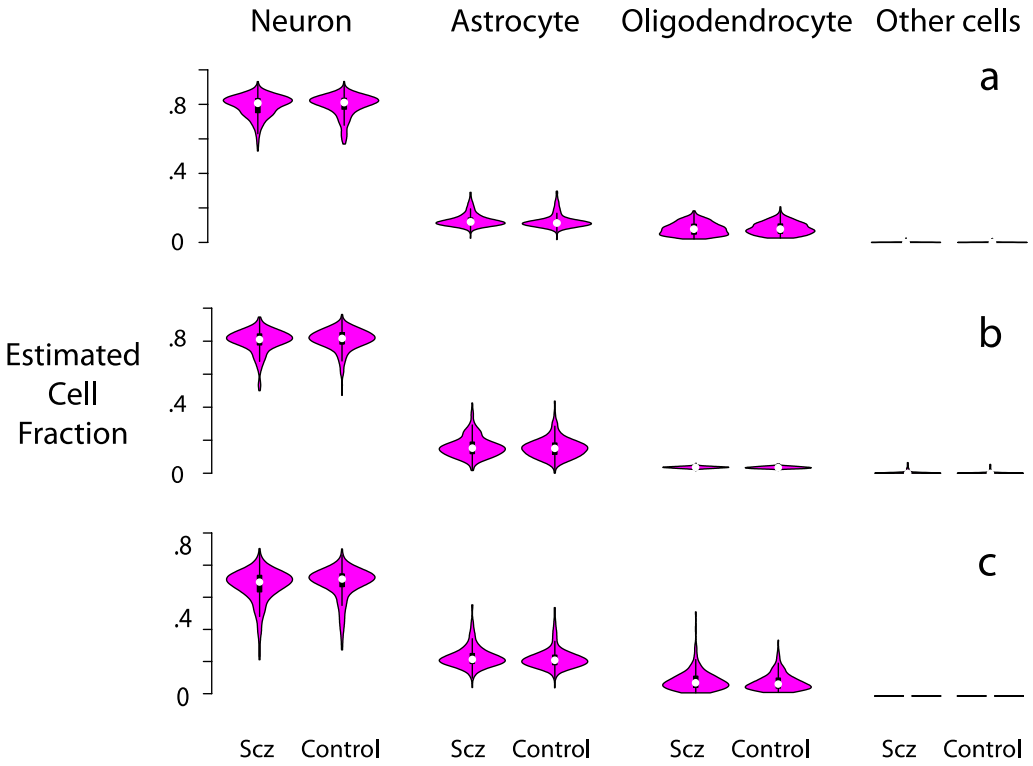


Panos Roussos
Robin Kramer

Additional analyses found in supplement

- Evaluation of 5’ to 3’ RNA read coverage using mRIN (Feng et al 2015)
- Technical validation by qPCR for select genes
- Isoform level analyses
- Cross-validation using 80/20 splits of the data for differential expression and WGCNA
- Effects of age on differential expression
- Effects of medication on differential expression including unpublished RNA sequencing data from mice and macaques treated with haldol
- Cell Type Deconvolution
- Correlation of CMC data with two prior meta-analytic studies
- Theoretical treatment of expectation with respect to effect likely to result from allele frequency differences observed between cases and controls

Cell fractions appear balanced between cases and controls



CIBERSORT to deconvolve the mixture based on 11,992 genes from the Zhang matrix of mouse cell-specific expression

CIBERSORT to deconvolve the mixture based on 415 human cell-specific markers (markers estimated from the data),

CellMix package, lsfit option, using 11,992 genes from the Zhang matrix

Bert Klei
Bernie Devlin
Andrew Browne
Menachem Fromer

Cell Marker lists from Zhang et al J Neurosci. 2014 34:11929-47
Analyses using CIBERSORT Newman et al. Nature Methods 12:453

Co-expression networks

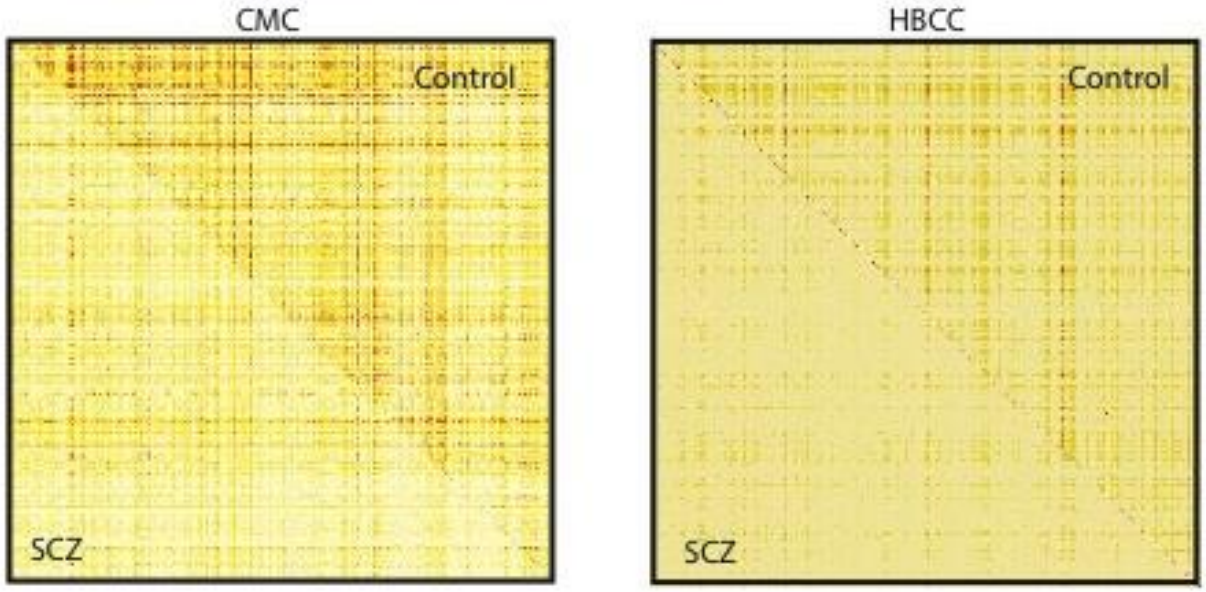
Exploring gene expression by weighted gene co-expression network analysis (WGCNA)

Langfelder and Horvath, 2008

1. Define correlation between all gene pairs.

2. Cluster into modules.

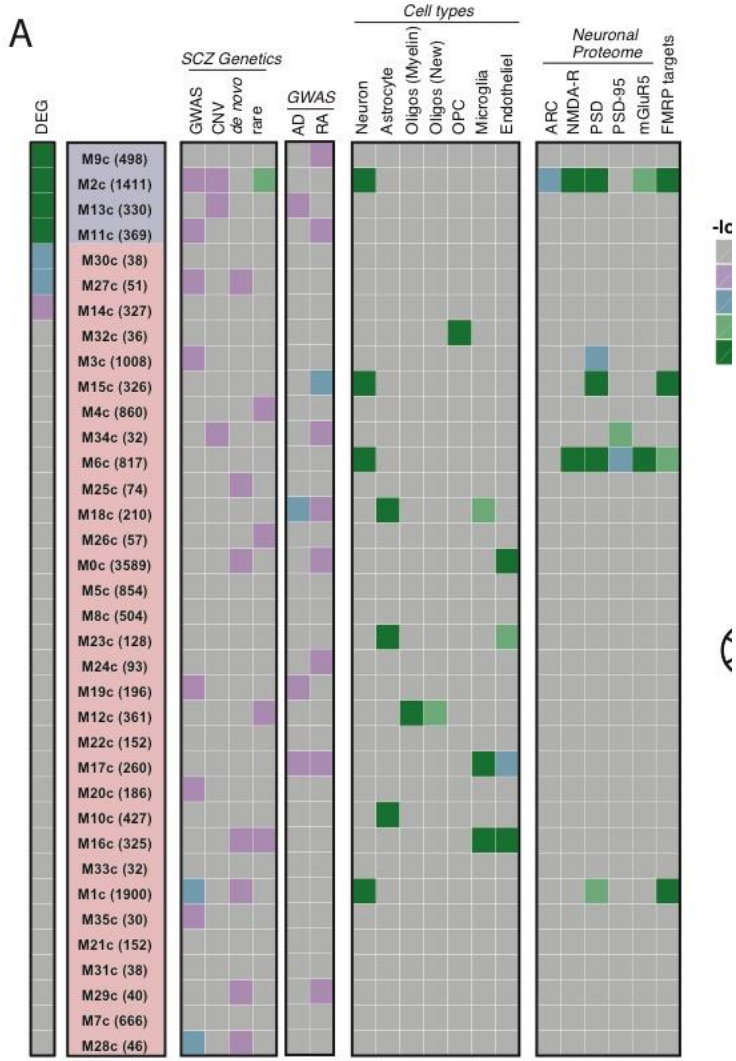
Differences?
Cell Types?
Genetics?



Off diagonal = Between modules

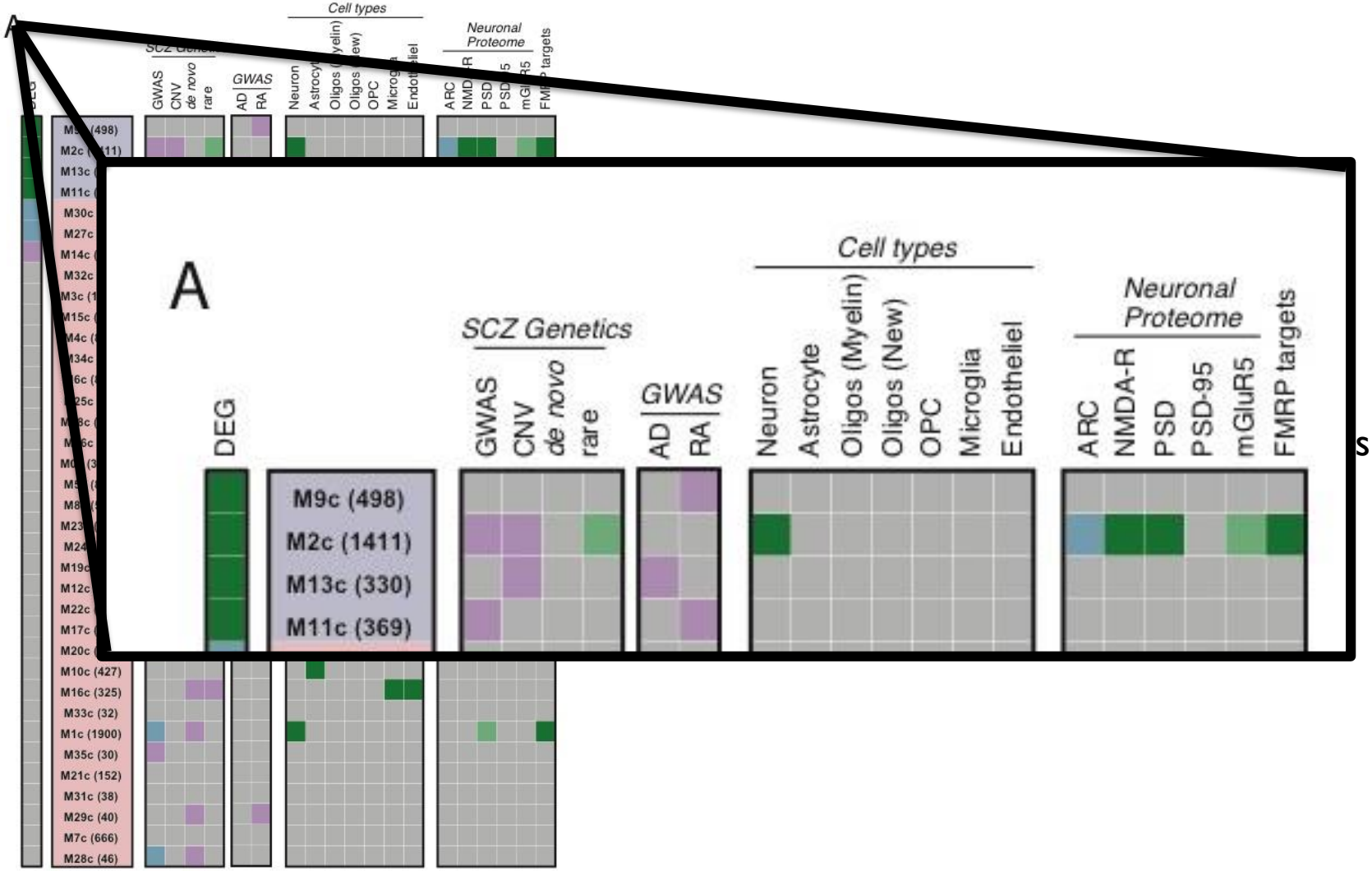
Panos Roussos

A single module of ~1400 genes that is most relevant

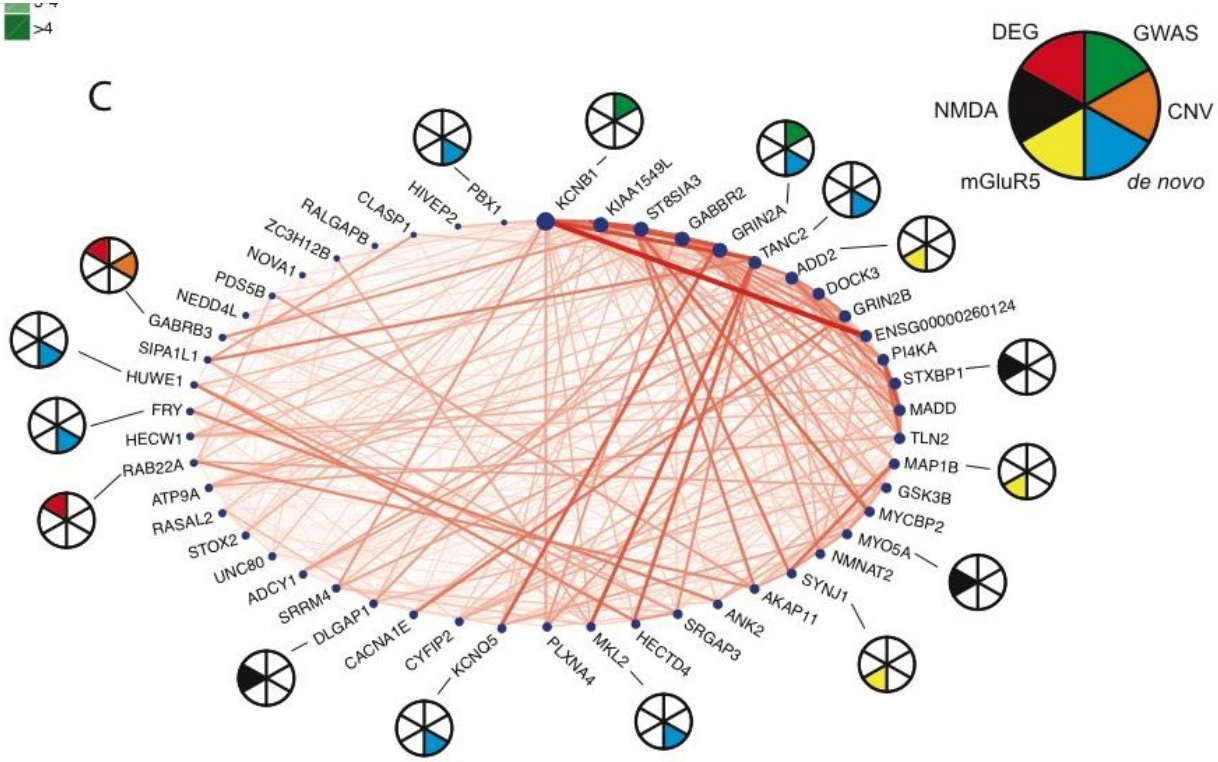


35 modules with of 30 to 1900 genes

A single module of ~1400 genes that is most relevant



Genes and connections

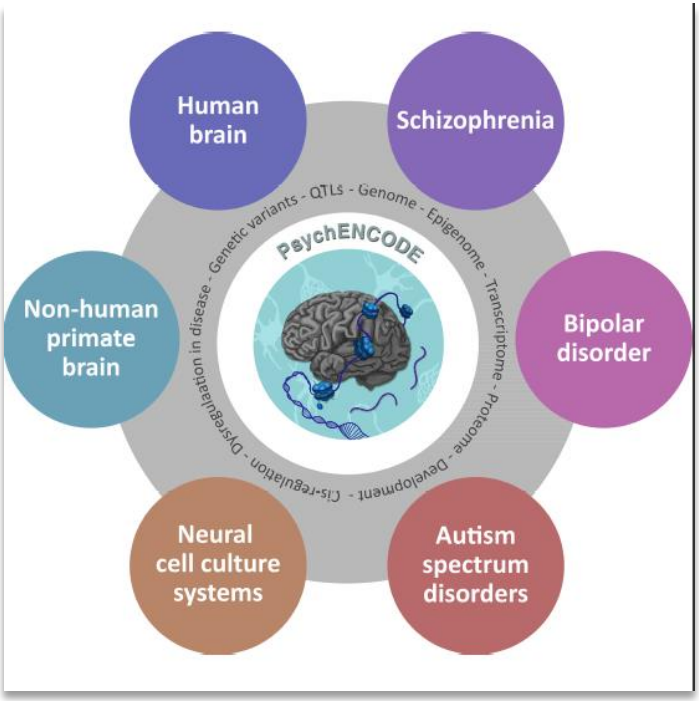


M2c module is enriched for multiple categories, including axon guidance, postsynaptic membrane, transmission across chemical synapses, and voltage-gated potassium channel complexes

PsychENCODE

PsychEncode Consortium

Nature Neuroscience 2015 18:1707



PRINCIPAL INVESTIGATORS

Schahram Akbarian
Gregory Crawford
Stella Dracheva
Peggy Farnham
Mark Gerstein
Daniel Geschwind
Andrew Jaffe
James Knowles
Chunyu Liu
Jonathan Mill
Angus Nairn
Mette Peters
Dalila Pinto
Nenad Sestan
Pamela Sklar
Matthew State
Patrick Sullivan
Flora Vaccarino
Sherman Weissman
Zhipeng Weng
Kevin White

RNAseq
ChipSeq
Open chromatin (ATAC-seq)
Chromosomal loop mapping
Methylation by bisulfite sequencing and NOMe-seq
Microwesterns

~1,000 samples

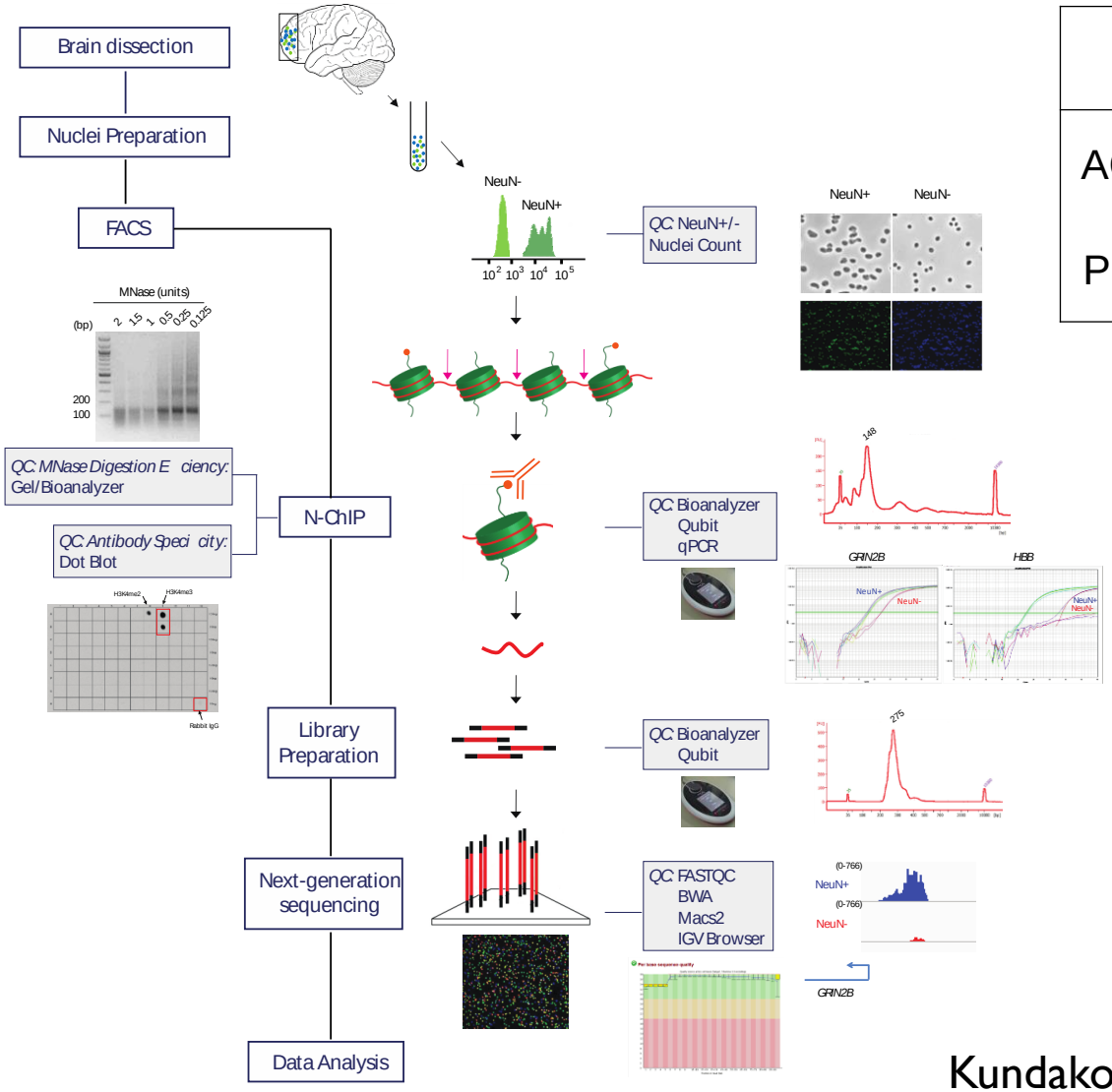
First data released January 2016

www.synapse.org/pec

https://www.synapse.org/#!Synapse:syn4921369

EpiMap Project

NIMH Human Brain Collection Samples



		H3K4me3	H3K27ac
ACC	NeuN+	15	16
	NeuN-	15	16
PFC	NeuN+	16	16
	NeuN-	17	16

Total N = 129

Goal: 80M paired ends
H3K27ac
40M for H3K4me3

Encode metrics:
NRF = 0.8
PBC = 0.8
NSC = 1.1/RSC = 0.8

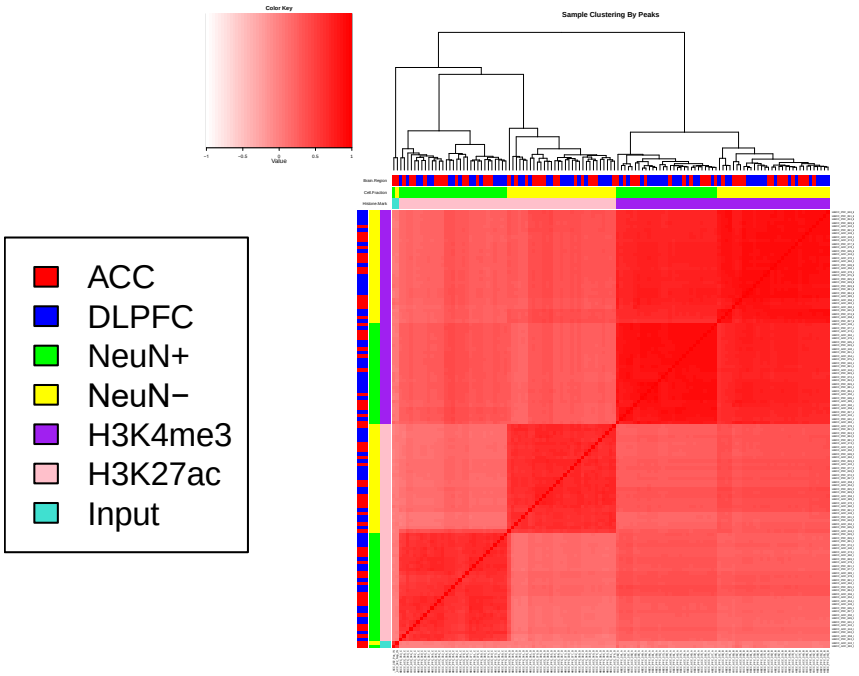
Kundakovic et al. Biol Psych March 9 2016

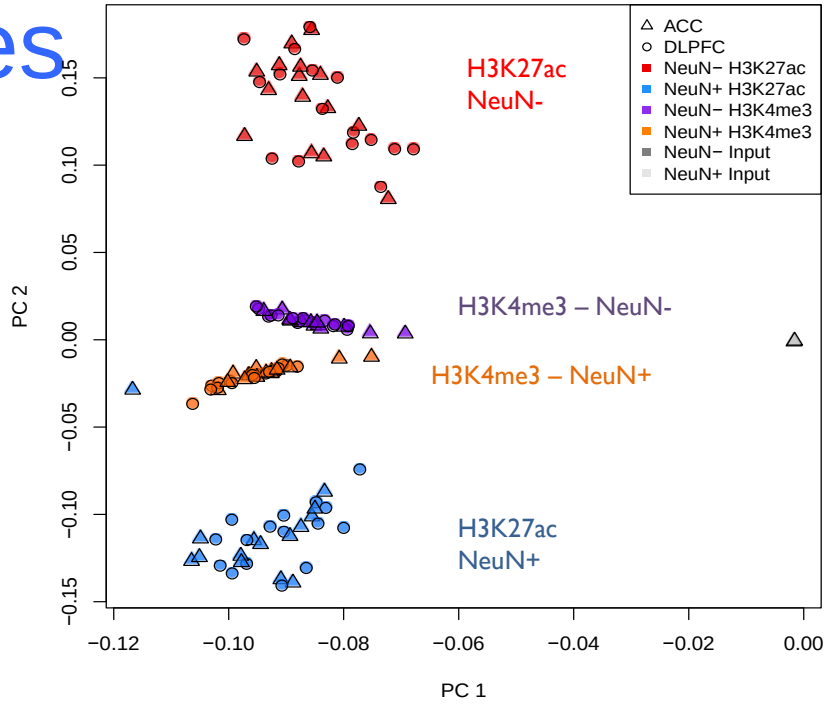
samples:

A chance to explore interindividual and cell-type differences in epigenetic profiles

40,000 peaks H3K4me3
~1.6% found only in neurons

100,000 peaks H3K27ac
~8% found only in neurons





Davy Kavanagh
Yan Jiang
Marija Kandakovic
Menachem Fromer
Kiran Girdhar
Gabriel Hoffman
Schahram Akbarian

Conclusions

- CommonMind: building **large resources** of molecular data from human brains

- eQTL can provide **functional clues** as to how SCZ risk variants mediate risk
  - 5 strong candidates for causal genes within GWAS loci, 3 of which have a brain phenotype in zebrafish

- Consistent with genetic polygenicity, **many genes are slightly disrupted in DLPFC for SCZ**

Acknowledgements

- Icahn School of Medicine at Mount Sinai
  - Division of Psychiatric Genomics:
    - **Panos Roussos**, Alexander Charney, Doug Ruderfer, Eli Stahl, Shaun Purcell, **Menachem Fromer, Davy Kavanagh**, Jessica Johnson, Kristen Brennand, Kiran Girdhar, Gabriel Hoffman
  - Eric Schadt, Ben Readhead, Bin Zhang, Jun Zhu
  - Schahram Akbarian, Yan Ziang, Marija Kandakovic, Amanda Mitchell, Venu Pothula, John Fullard
- CommonMind Consortium
  - **Bernie Devlin, Bert Klei**, Enrico Dominici, **Solly Sieberts**, Kristen Dang, Thaneer Perumal, Mette Peters
- Duke/UNC- Nico Katsanis, Ed Oh, Mahsa Parvizi, Pat Sullivan
- PsychCHIP – Ben Neale, Jennifer Moran and many others
- BD/SCZ group PGC-Doug Ruderfer, Ken Kendler, and others
- Psychiatric Genomic Consortium, PsychENCODE Consortium
- Public and other Datasets
  - Epigenomic Data: ENCODE and Roadmap to Epigenomics Projects
  - Gene expression: BrainCloud,NIA/NIH and Harvard Brain Tissue Resource Center