Big Data and Genetic Liability to Neuropsychiatric Disease

FBI Cyber Series
May 21st, 2020

Panos Roussos, M.D., Ph.D.

The Pamela Sklar Division of Psychiatric Genomics
Icahn Institute for Data Science and Genomic Technology
Friedman Brain Institute
Department of Genetics and Genomic Sciences
Department of Psychiatry
James J. Peters VA Medical Center
Mental Health Facts in America

Fact: 43.8 million adults experience mental illness in a given year.

- 1 in 3 adults in America experience a mental illness.
- Nearly 1 in 25 (10 million) adults in America live with a serious mental illness.
- One-half of all chronic mental illness begins by the age of 14; three-quarters by the age of 24.

Prevalence of Mental Illness by Diagnosis

- 1.1% (1 in 100 (2.4 million) American adults live with schizophrenia.
- 2.6% (6.1 million) of American adults live with bipolar disorder.
- 6.9% (16 million) of American adults live with major depression.
- 18.1% (47 million) of American adults live with anxiety disorders.

47M Worldwide

5.7M Americans

6th Leading cause of death
Neuropsychiatric disorders are the leading cause of disease burden in the U.S.

**Age standardized disability adjusted life years (DALY) rate per 100,000 population in both sexes**

<table>
<thead>
<tr>
<th>Disease Category</th>
<th>2015</th>
<th>1990</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neuropsychiatric disorders</strong></td>
<td>3355</td>
<td>2779</td>
</tr>
<tr>
<td><strong>Cancer</strong></td>
<td>3131</td>
<td>4050</td>
</tr>
<tr>
<td><strong>Cardiovascular diseases</strong></td>
<td>3065</td>
<td>5070</td>
</tr>
<tr>
<td><strong>Injuries</strong></td>
<td>2419</td>
<td>3390</td>
</tr>
<tr>
<td><strong>Musculoskeletal disorders</strong></td>
<td>2357</td>
<td>2442</td>
</tr>
<tr>
<td><strong>Chronic respiratory diseases</strong></td>
<td>1050</td>
<td>1103</td>
</tr>
<tr>
<td><strong>Endocrine</strong></td>
<td>923</td>
<td>854</td>
</tr>
<tr>
<td><strong>Skin diseases</strong></td>
<td>642</td>
<td>640</td>
</tr>
</tbody>
</table>
Neuropsychiatric disorders are highly heritable

**Lifetime prevalence**

- ANX
- AAD
- MDD
- ADHD
- PTSD
- BPD
- EAT
- OCD
- ASD
- SCZ
- TS

**Twin/family based heritability**

- ANX
- AAD
- MDD
- ADHD
- PTSD
- BPD
- EAT
- OCD
- ASD
- SCZ
- TS
Genome-wide association study to associate genetic variations with traits

PGC, Nature 2014
GWAS has identified hundreds of risk loci.
Challenges in the post-GWAS era
Where to go next?
What are the associated loci?
What are the causal variants?
What are the causal genes?
What is the causal mechanisms?
What is the tissue of action?
What structure of the 3D genome is perturbed?

Gene Expression

Liver

Brain
What regulatory network is perturbed?
Outline

1. Linking risk variants to changes in gene expression
2. Identifying vulnerable cellular populations
3. Studying disease associated change in the 3D genome
4. Performing multiscale omics integration with risk variation
5. Defining AD-associated microglia states and subpopulations
Functional variants are more likely to be associated with a trait

Roussos, et al, Cell Reports 2014
Building large resources to study gene expression in the human brain

CommonMind Consortium (CMC)
Includes >1,000 cases with Schizophrenia and bipolar disorder and controls

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

Haroutounian, Lewis, Gur, Hahn, Lipska
Gene expression is only subtly disrupted


No. of genes upregulated: 332
No. of genes downregulated: 361
Identification of a disease-associated subnetwork

Linking risk variation to gene expression changes

- Genotypes
- Gene expression
- GReX
- Gene-trait association studies
- RNAseq/Genotype
- GWAS

Gene expression data

RNA

DNA

SNPs

Genotype data
RESULTS: Highlighted 37 genes in 20 GWAS loci (Bonferroni-Sherlock \( p < 0.05 \))

DATA: 12,367 Ensembl genes and PGC SCZ2 GWAS data

20% of SCZ loci may increase risk by altering gene expression

5 LOCI WHERE ONLY ONE GENE IS IMPLICATED IN THE GWAS LOCUS

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

Functional validation of SCZ-associated genes

Underexpression and overexpression lead to decreased head size

Limited power to detect gene expression changes driven by risk variants

Genome-wide, the median number of subjects with SCZ and controls needed to obtain 80% power assuming 10,000 genes is \( \approx 28,500 \)
Risk variants affect the distal regulation of gene expression

- 24 RNA-seq based eQTL studies from CMC, STARNET and GTEx
- GWAS for 57 traits

60% of associated genes to risk variants are not mapped to the closest gene
### Risk variants show tissue specificity

**Nat Comm 2019**

![Heatmap image with tissue specificity data](image-url)

**Wen Zhang**

**Georgios Voloudakis**
Relevant tissues identify more unique genes
An example at the gene level

<table>
<thead>
<tr>
<th>Gene</th>
<th>CMC Brain, DLPFC</th>
<th>STARNET Artery, aorta</th>
<th>STARNET Mammary, artery</th>
<th>STARNET Adipose, subcutaneous</th>
<th>STARNET Adipose, visceral</th>
<th>STARNET Liver</th>
<th>GTEx Artery, aorta</th>
<th>GTEx Adipose, subcutaneous</th>
<th>GTEx Adipose, visceral</th>
<th>GTEx Liver</th>
<th>GTEx Muscle, skeletal</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMDT1</td>
<td># # # # # # #</td>
<td># # # # #</td>
<td>#</td>
<td></td>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMG6</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td></td>
<td></td>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNAP91</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNORD3B-2</td>
<td>#</td>
<td></td>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNX19</td>
<td>#</td>
<td></td>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOWAHC</td>
<td>#</td>
<td></td>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Legend

- Significant association (FDR < 0.05)
- Significant P values (<0.05)
Limitations of gene expression colocalization studies

Correlated expression across individuals may cause false hits

Wainberg et al, Nat Gen 2019

No insights into the putative regulatory mechanisms

Aggregate score at the gene level without considering the complexity of transcriptome diversity
Epigenome profiling in human brain tissue

Nuclei from fresh frozen brain tissue

Non-neuronal

Neuronal

HiC and Capture-C

H3K4me3 (promoters)

H3K27ac (enhancers)

ATACseq (open chromatin)
SCZ risk loci are enriched within neuronal enhancers

To understand if histone e.g. H3K27ac, H3K4me3 modifications can provide any quantifiable measures of heritable predisposition to Schizophrenia.
SCZ risk loci are enriched within neuronal open chromatin regions

To understand if open chromatin in enhancer and promoters can provide any quantifiable measures of heritable predisposition to Schizophrenia.

Collaborative effort with the Hurd and Ehrlich Labs
SCZ risk loci are enriched within glutamate neuron open chromatin regions

Brain_region_abbreviation
- ACC
- DLPFC
- PVC

Cell Type
- GLU
- GABA
- OLG
- MGAS

Collaborative effort with the Dracheva and Ehrlich Labs

Under Revision
AD risk loci are enriched within microglia open chromatin regions

Fresh Tissue x250 samples

CD45+/CD11b+ FACS

Multiscale omics and single cell omics

Collaborative effort with the Haroutunian, Charney, Kellner and Bennett Labs
Large-scale epigenetic studies in SCZ and AD

SCZ (CommonMind Consortium)

Collaborative effort with the Akbarian Lab

AD (AMP-AD)

Collaborative effort with the Haroutunian Lab

In Preparation
Epigenetic regions are perturbed in neuropsychiatric traits

SCZ (CommonMind Consortium)

AD (AMP-AD)

<table>
<thead>
<tr>
<th>Trait associations</th>
<th>EC</th>
<th>STG</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fractions of OCRs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATAC Neuron</td>
<td>149</td>
<td>1795</td>
<td>12282</td>
</tr>
<tr>
<td>ATAC non-Neuron</td>
<td>0</td>
<td>5</td>
<td>6165</td>
</tr>
<tr>
<td>All</td>
<td>476</td>
<td>4406</td>
<td>33386</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Dx</th>
<th>BBScore</th>
<th>PlaqueMean</th>
<th>CDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>3</td>
<td>1</td>
<td>281</td>
<td>534</td>
</tr>
<tr>
<td>STG</td>
<td>5</td>
<td>12</td>
<td>0</td>
<td>4122</td>
</tr>
<tr>
<td>All</td>
<td>31</td>
<td>435</td>
<td>2085</td>
<td>9514</td>
</tr>
</tbody>
</table>
Perturbed epigenetic regions are enriched with risk variation

SCZ (CommonMind Consortium)

AD (AMP-AD)
Studying the 3D genome

HiC (or 3C capture techniques)

Correlation structure of epigenome data (ATACseq and ChIPseq) based on population variation

Correlation structure of epigenome data (scATACseq) based on cell variation

Rowley & Corces Nat Rev Genet 2018

Delaneau et al Science 2019
Hoffman et al Bioinformatics 2020

Pliner et al Mol Cell 2018
Correlation structure of epigenome data to define CRDs

1. Compute location correlations
   - Fast calculation of correlations

2. Hierarchical clustering of features
   - Clustering uses spatial ordering

3. Produce discrete clusters of features
   - Define clusters at multiple resolutions

4. Filter, combine clusters
   - Remove weak and redundant clusters

5. Statistical test of differential correlation
   - Choose from multiple tests to compare correlation matrices

6. Data Visualization
   - Plot genes, feature locations, clusters, correlations

CRD = chromatin regulatory domain

Bioinformatics 2020
https://github.com/GabrielHoffman/decorate
CRDs capture 3D chromatin organization
CRDs capture 3D structure with increased resolution

SCZ (CommonMind Consortium)

AD (AMP-AD)
CRD borders are enriched in insulators and CTCF binding sites

**SCZ (CommonMind Consortium)**

- H3K4me3-H3K4me3
- H3K27ac-H3K27ac
- HIC TAD in PFC Homog.

**AD (AMP-AD)**

- neuron_BM22
- neuron_BM36
- glia_BM22
- glia_BM36
- NeuN TAD 20Kb
- Glia TAD20Kb
Disease-associated CRDs converge to molecular pathways

SCZ (CommonMind Consortium)
- Voltage-gated Potassium channels
- GABA receptor activation
- Tyrosine Kinase signaling
- NOTCH3 signaling

AD (AMP-AD)
- Calmodulin induced events
- DAG and IP3 signaling
- HDAC deacetylation
- WNT signaling
Multiscale FUN-WAS using PolyXcan in neuropsychiatric traits

- ATAC-seq
  - DLPFC homogenate (n=164)
  - Neuron BM22 (n=108)
  - Neuron BM36 (n=97)
  - Glia BM22 (n=106)
  - Glia BM36 (n=94)

- ChIP-seq
  - HBCC H3K27ac (n=122)
  - CMC H3K27ac (n=191)
  - CMC H3K4me3 (n=163)

- Chrom. accessibility
  - Neuron
  - Glia

- RNA-seq
  - GReX
  - Genes (Capstone, n=924) Used when no isoform information
  - Isoforms (Capstone, n=924) Used when isoform-level information

37 neuropsychiatric traits
4 control traits (HbA1c, DMII, Crohn’s disease and ulcerative colitis)

In Preparation
PolyXcan increases power to detect gene-traits associations

KS p value < $2.22 \times 10^{-16}$

PolyXcan performance improvement: 22.9 times in effective sample size for common predictions of gene-trait associations (10,569 genes)
A few examples: ZNF823 association in SCZ

Transcript z-score = 4.35
Unit z-score = 28.4
A few examples: DRD2 association in SCZ

Transcript z-score = NA
Unit z-score = 15.1
Exploring the heterogeneity of brain immune cells in AD

N = 30 (Control: 14, Other: 5, AD: 9, PD: 2)
Dissection from prefrontal cortex (Biopsy: 14, autopsy: 16)
Mechanical/Enzymatic Dissociation
CD45⁺ /CD11b⁺ FACS
scRNA-seq (10x)

In Preparation

Yungil Kim
John Fullard
De novo taxonomy in the human brain immune cells
Association of degenerative microglia with AD

Stage 0: C1, C7, C8
Stage I, II: C3, C6, C11
Stage III: C5, C10

|Cell|^{(4)}APS+(9)^CC

AD progression stage (APS)

Cluster composition (CC)

Pearson residuals

Alzheimer's Disease
Annotating brain T cells and monocytes
Brain CD8+ cells are more prevalent in AD cases.

**Article**

Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer’s disease.

https://doi.org/10.1038/s41586-019-1895-7

Received: 30 May 2018

Accepted: 2 December 2019

Published online: 8 January 2020
Take home messages

1. Higher order chromatin structure and epigenomic features are perturbed in SCZ and AD

2. Application of machine learning approaches for integrative omics analysis has the promise to help in the functional characterization of risk variation across neuropsychiatric traits

3. AD is associated with perturbation in microglia states and other peripheral immune cells