

**Interpretable deep learning model,
embedding the gene regulatory network,
for understanding functional genomics in neuropsychiatric disorders**



M Gerstein
Yale
(See last slide for more info.)

Slides freely downloadable
from Lectures.GersteinLab.org
& “tweetable” (via [@MarkGerstein](https://twitter.com/MarkGerstein))

Sample Sources: >2,500 brains

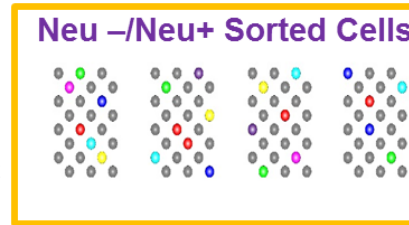
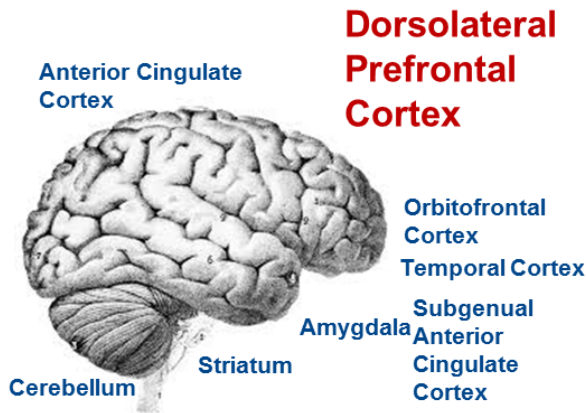
Genome:
WGS, genotype

Epigenome:
ChIP-seq, ATAC-seq, HiC, ERRBS, Array Methylation, NOMeSeq

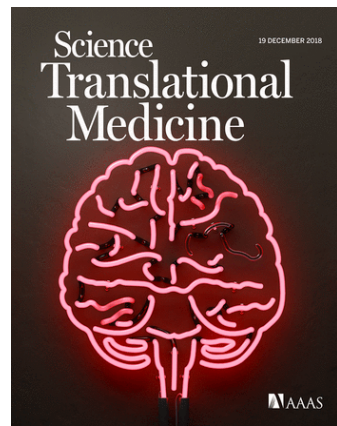
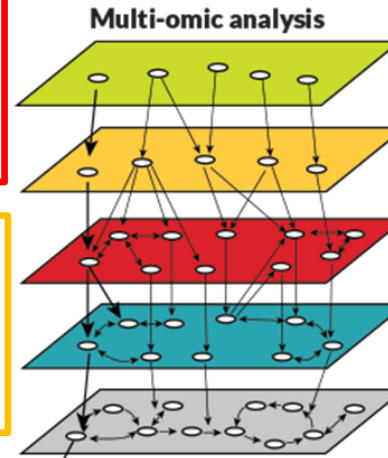
Transcriptome:
RNA-seq, lncRNAseq,

Proteome:
MWP, LC-MS/MS

Cross-disorder: ASD, SCZ, BP, Neurodevelopmental, Neurotypical



Single Cell



PsychENCODE

'18 rollout in Science

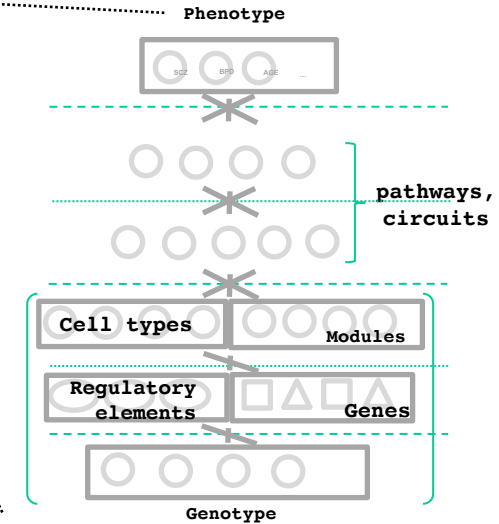
11 papers in total.

Major material in the 3 capstones:

Wang et al. ('18), Li et al. ('18), Gandal et al. ('18)

A core issue addressed by PsychENCODE: Using functional genomics to reveal molecular mechanisms between genotype and phenotype in brain disorders

Disease	Heritability*	Molecular Mechanisms
Schizophrenia	81%	(C4A)
Bipolar disorder	70%	-
Alzheimer's disease	58 - 79%	Apolipoprotein E (APOE), Tau
Hypertension	30%	Renin–angiotensin–aldosterone
Heart disease	34-53%	Atherosclerosis, VCAM-1
Stroke	32%	Reactive oxygen species (ROS), Ischemia
Type-2 diabetes	26%	Insulin resistance
Breast Cancer	25-56%	BRCA, PTEN



Many psychiatric conditions are highly heritable

Schizophrenia: up to 80%

But we don't understand basic molecular mechanisms underpinning this association

(in contrast to many other diseases such as cancer & heart disease)

Thus, interested in developing predictive models of psychiatric traits which:

Use observations at intermediate (molecular levels) levels to inform latent structure

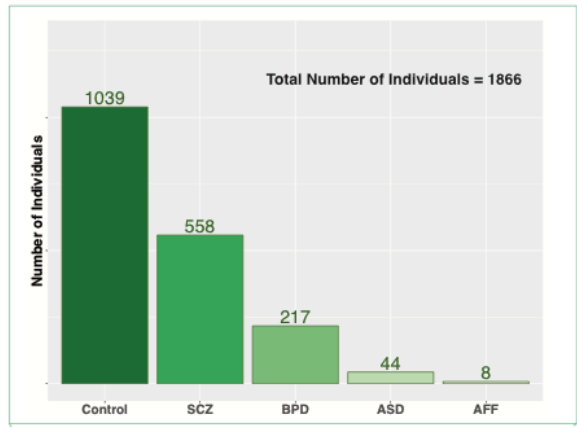
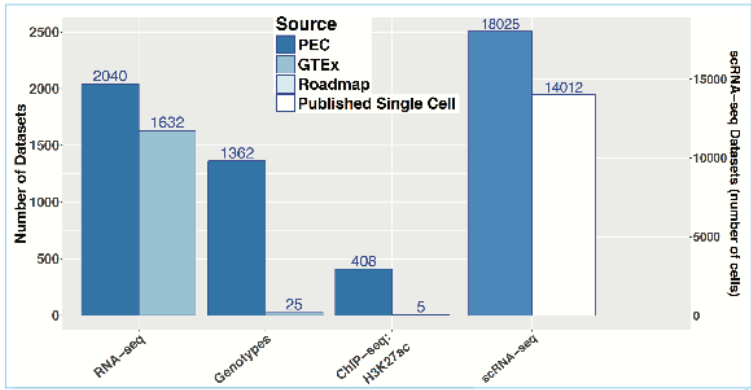
Use the predictive features of these “molecular endo phenotypes” to begin to suggest actors involved in mechanism

Interpretable deep learning model, embedding the gene regulatory network, for understanding functional genomics in neuropsychiatric disorders

- Construction of an adult brain resource with 1866 individuals
- Using the changing proportions of cell types (via **single-cell deconvolution**) to account for expression variation across a population
- Large-scale processing defines ~79K PFC **enhancers & creates a comprehensive QTL** resource (~2.5M eQTLs + cQTLs & fQTLs)
- Connecting QTLs, enhancer activity relationships & Hi-C contacts into a **brain regulatory network** & using this to link SCZ GWAS SNPs to genes
- Embedding the reg. network in a **deep-learning model** to predict psychiatric disease from genotype & transcriptome. Using this to suggest specific pathways & genes, as potential drug targets.
- Other resource uses: highlighting aging related genes

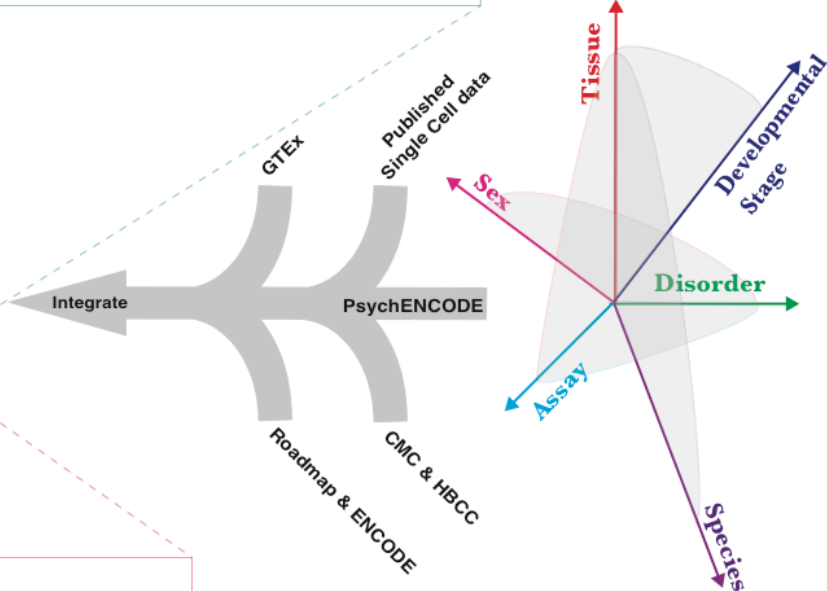
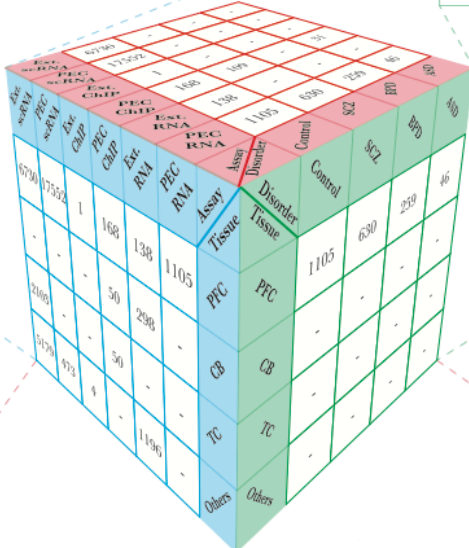
Interpretable deep learning model, embedding the gene regulatory network, for understanding functional genomics in neuropsychiatric disorders

- Construction of an adult brain resource with 1866 individuals
- Using the changing proportions of cell types (via **single-cell deconvolution**) to account for expression variation across a population
- Large-scale processing defines ~79K PFC **enhancers & creates a comprehensive QTL** resource (~2.5M eQTLs + cQTLs & fQTLs)
- Connecting QTLs, enhancer activity relationships & Hi-C contacts into a **brain regulatory network** & using this to link SCZ GWAS SNPs to genes
- Embedding the reg. network in a **deep-learning model** to predict psychiatric disease from genotype & transcriptome. Using this to suggest specific pathways & genes, as potential drug targets.
- Other resource uses: highlighting aging related genes



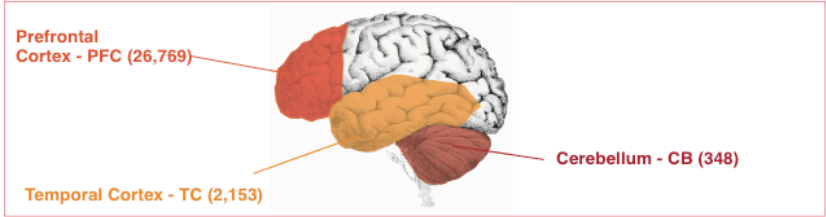
1866
 Individuals
 ~3.7K bulk RNA-seq
 ~32K single-cells

Disorder



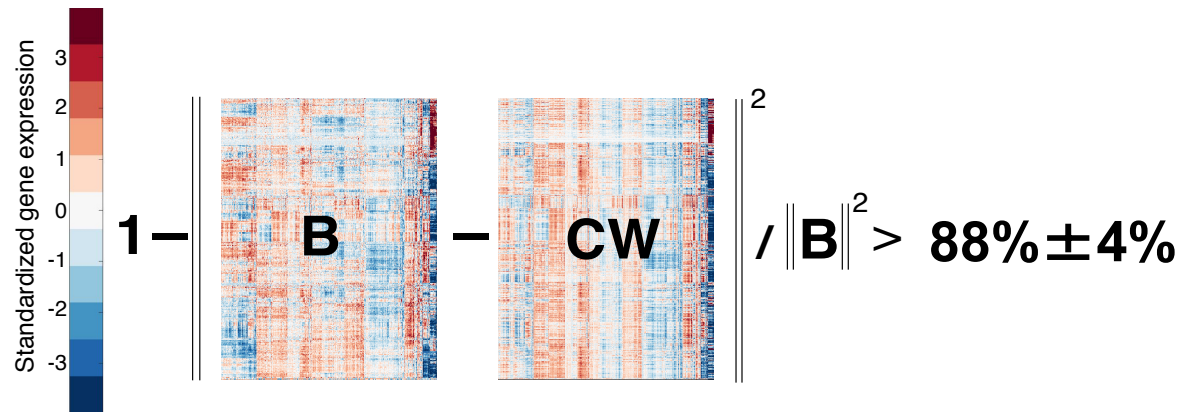
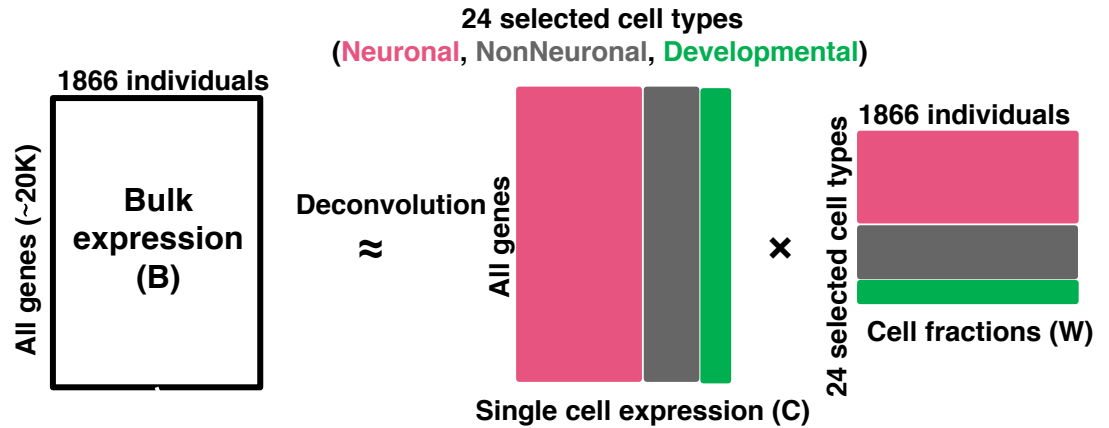
Collecting functional genomic datasets for the adult brain

from PsychENCODE, other large consortia & single cell studies



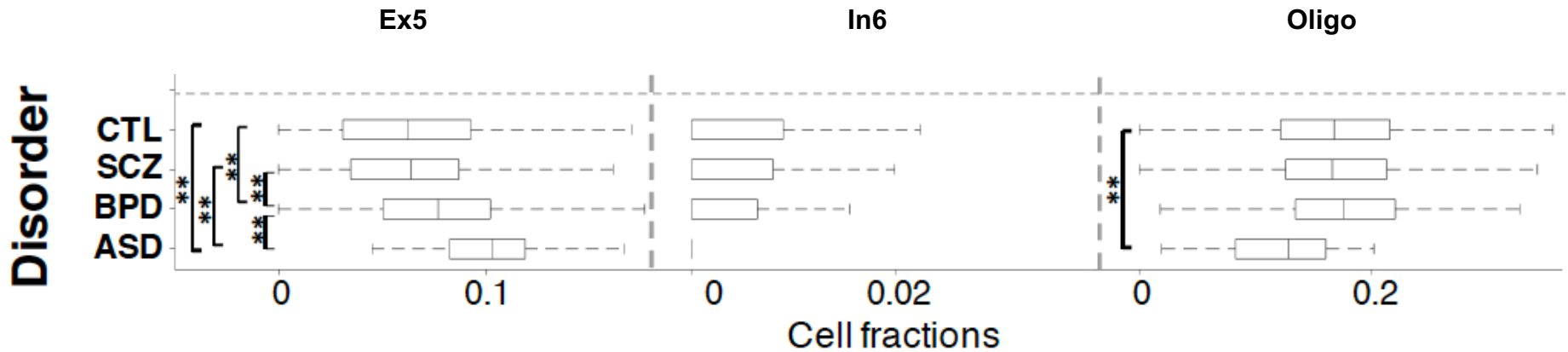
Single-cell deconvolution :

Supervised learning to estimate cell fractions



Individual and cross-population reconstruction accuracy via deconvolution

Different neuronal & glial cell fractions across disorders



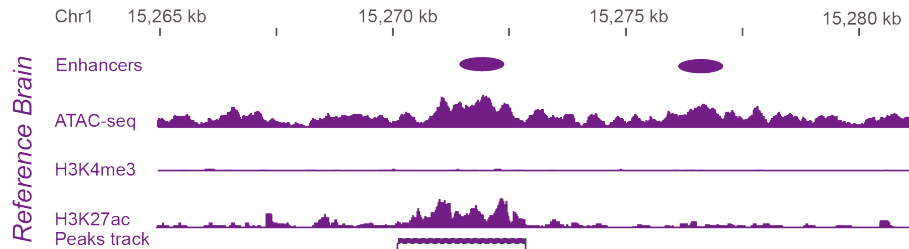
Excitatory to Inhibitory imbalance at neuronal subtype level for ASD*

* Rubenstein et al., Model of autism: increased ratio of excitation/inhibition in key neural systems, Genes Brain Behav. 2003

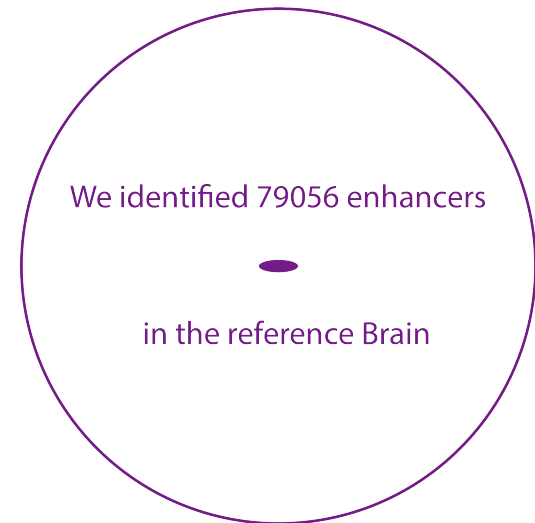
Interpretable deep learning model, embedding the gene regulatory network, for understanding functional genomics in neuropsychiatric disorders

- Construction of an adult brain resource with 1866 individuals
- Using the changing proportions of cell types (via **single-cell deconvolution**) to account for expression variation across a population
- Large-scale processing defines ~79K PFC **enhancers & creates a comprehensive QTL** resource (~2.5M eQTLs + cQTLs & fQTLs)
- Connecting QTLs, enhancer activity relationships & Hi-C contacts into a **brain regulatory network** & using this to link SCZ GWAS SNPs to genes
- Embedding the reg. network in a **deep-learning model** to predict psychiatric disease from genotype & transcriptome. Using this to suggest specific pathways & genes, as potential drug targets.
- Other resource uses: highlighting aging related genes

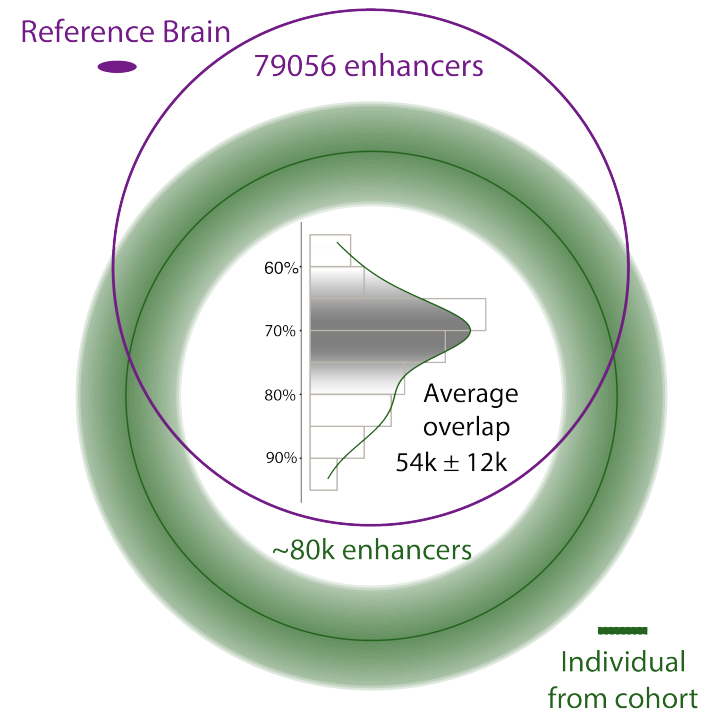
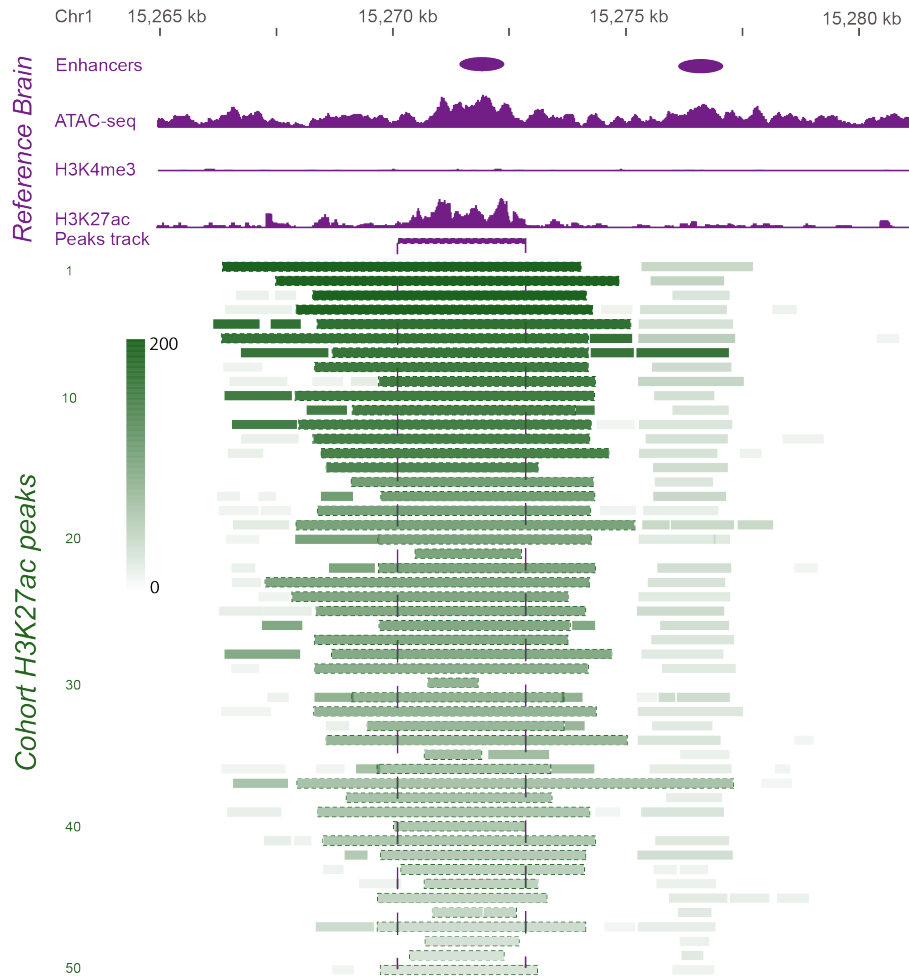
Developing a Reference Set of ~79K PFC Enhancers & Studying Their Population Variation



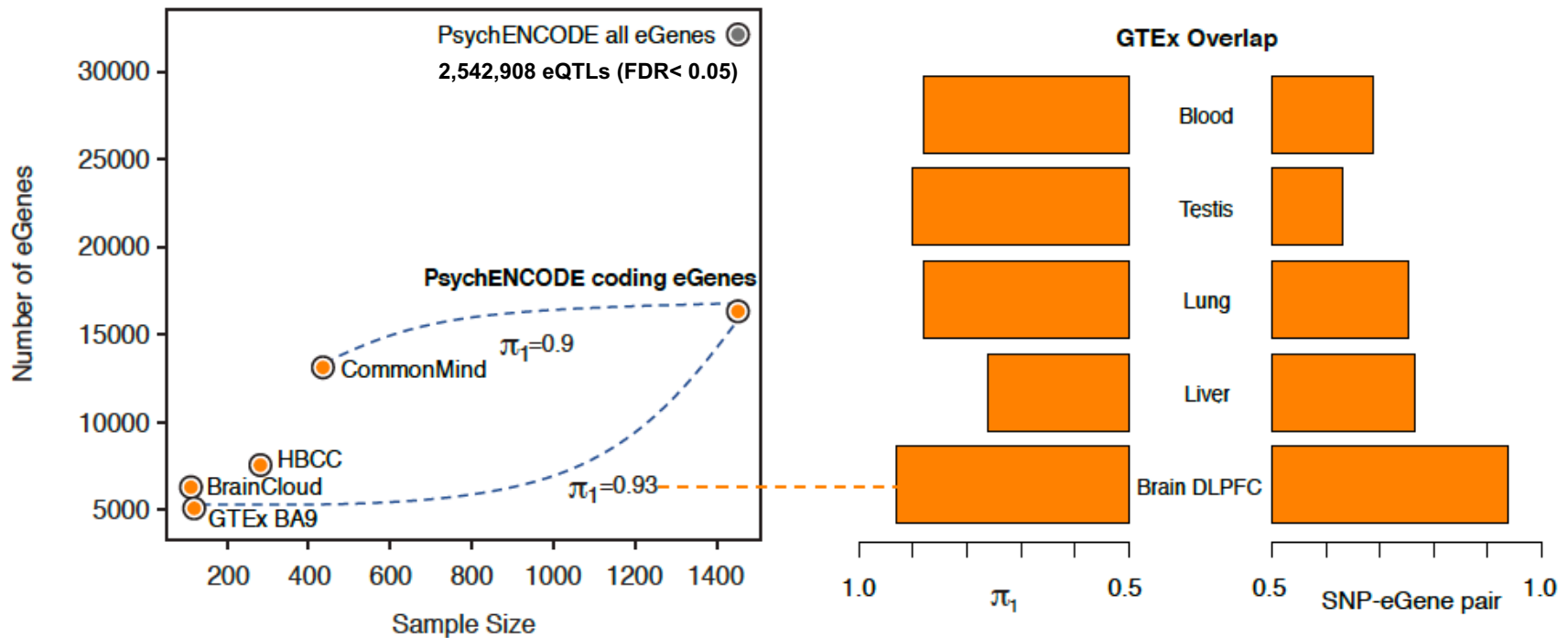
Consistent with ENCODE, active enhancers are identified as open chromatin regions enriched in H3K27ac and depleted in H3K4me3



Developing a Reference Set of ~79K PFC Enhancers & Studying Their Population Variation

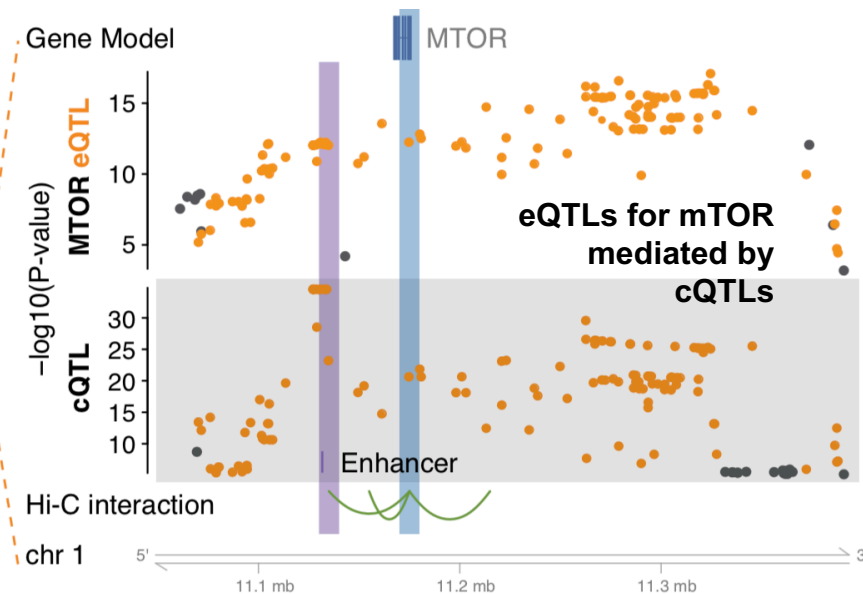
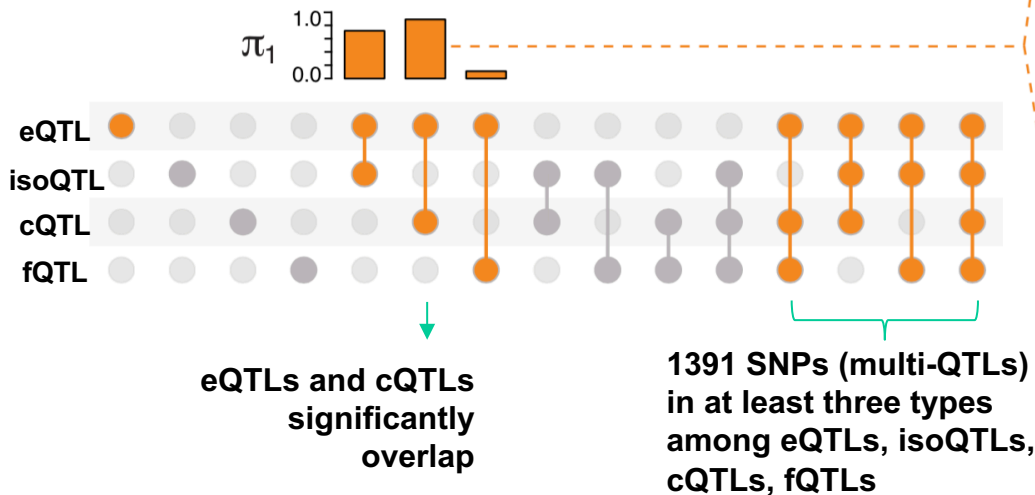


Larger brain eQTL sets than previous studies, but strong overlap with them



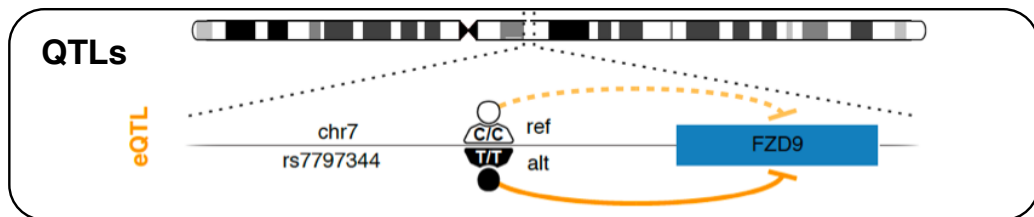
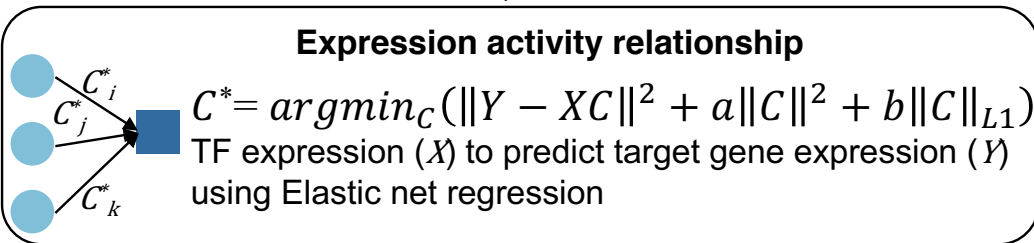
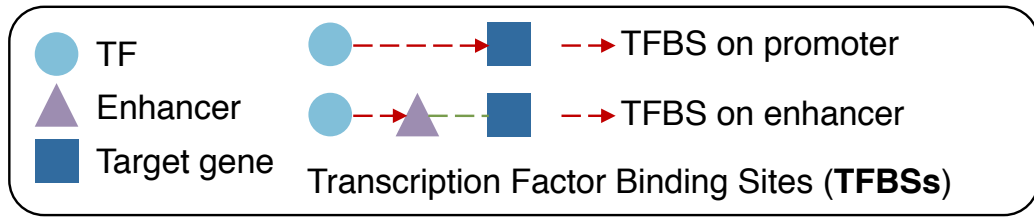
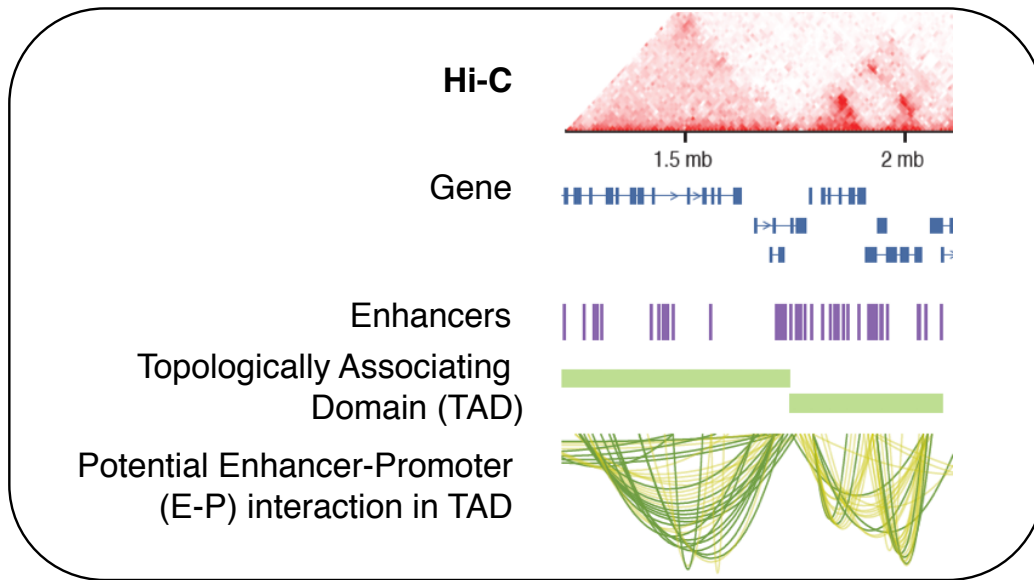
multi-QTLs from overlapping different types of QTLs: cQTL, fQTL, eQTL & isoQTL

	Numbers of QTLs	eGenes Enhancers Cell types	SNPs
eQTL	2,542,908	32,944	1,341,182
isoQTL	2,628,259	19,790	1,052,939
cQTL*	8,464	8,484	7,983
fQTL	4,199	9	1,672



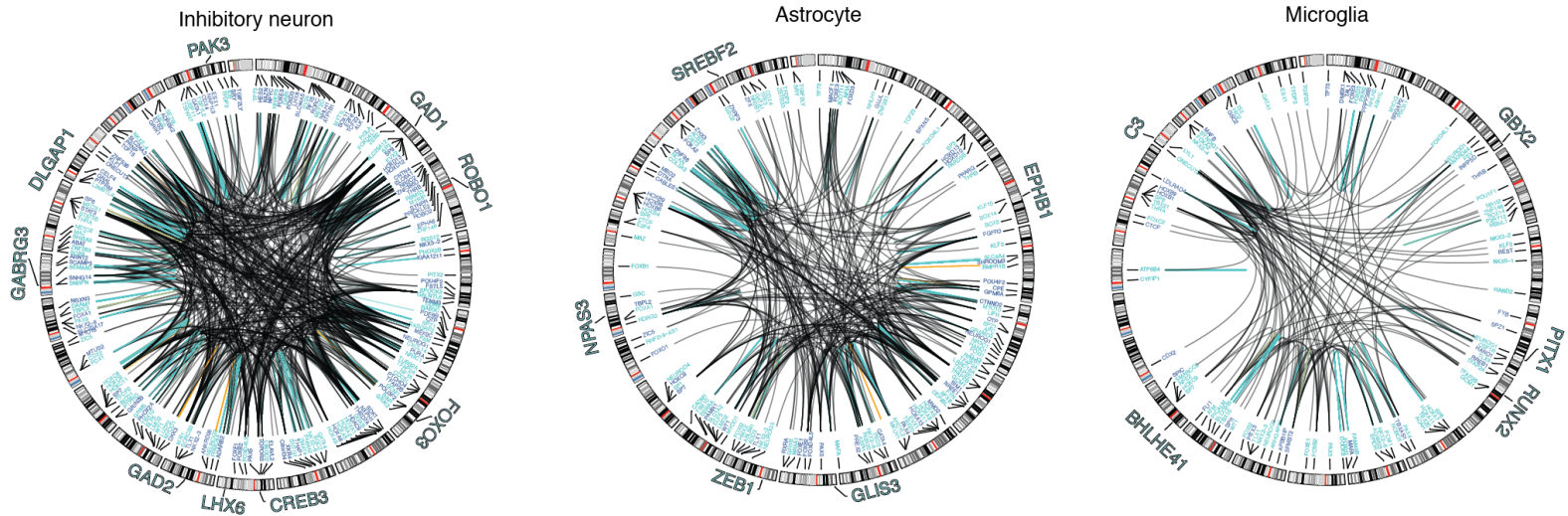
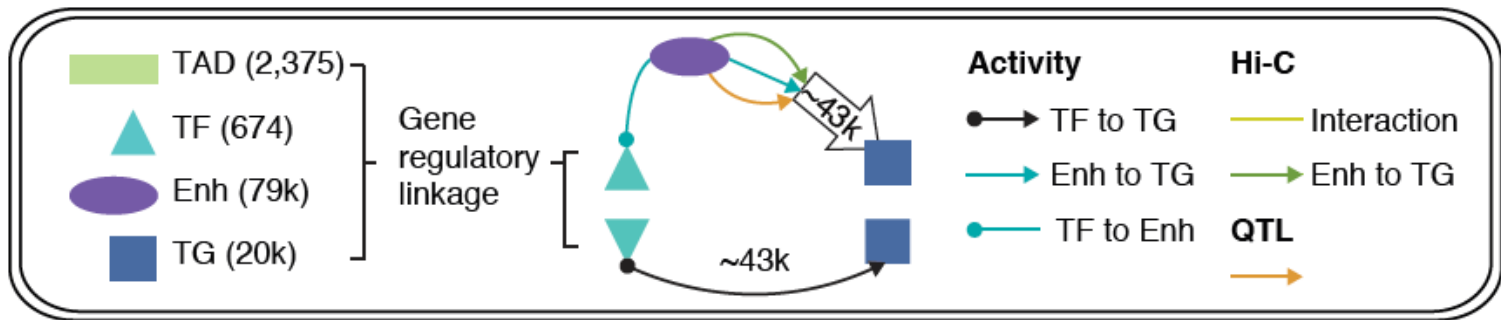
Interpretable deep learning model, embedding the gene regulatory network, for understanding functional genomics in neuropsychiatric disorders

- Construction of an adult brain resource with 1866 individuals
- Using the changing proportions of cell types (via **single-cell deconvolution**) to account for expression variation across a population
- Large-scale processing defines ~79K PFC **enhancers & creates a comprehensive QTL** resource (~2.5M eQTLs + cQTLs & fQTLs)
- Connecting QTLs, enhancer activity relationships & Hi-C contacts into a **brain regulatory network** & using this to link SCZ GWAS SNPs to genes
- Embedding the reg. network in a **deep-learning model** to predict psychiatric disease from genotype & transcriptome. Using this to suggest specific pathways & genes, as potential drug targets.
- Other resource uses: highlighting aging related genes



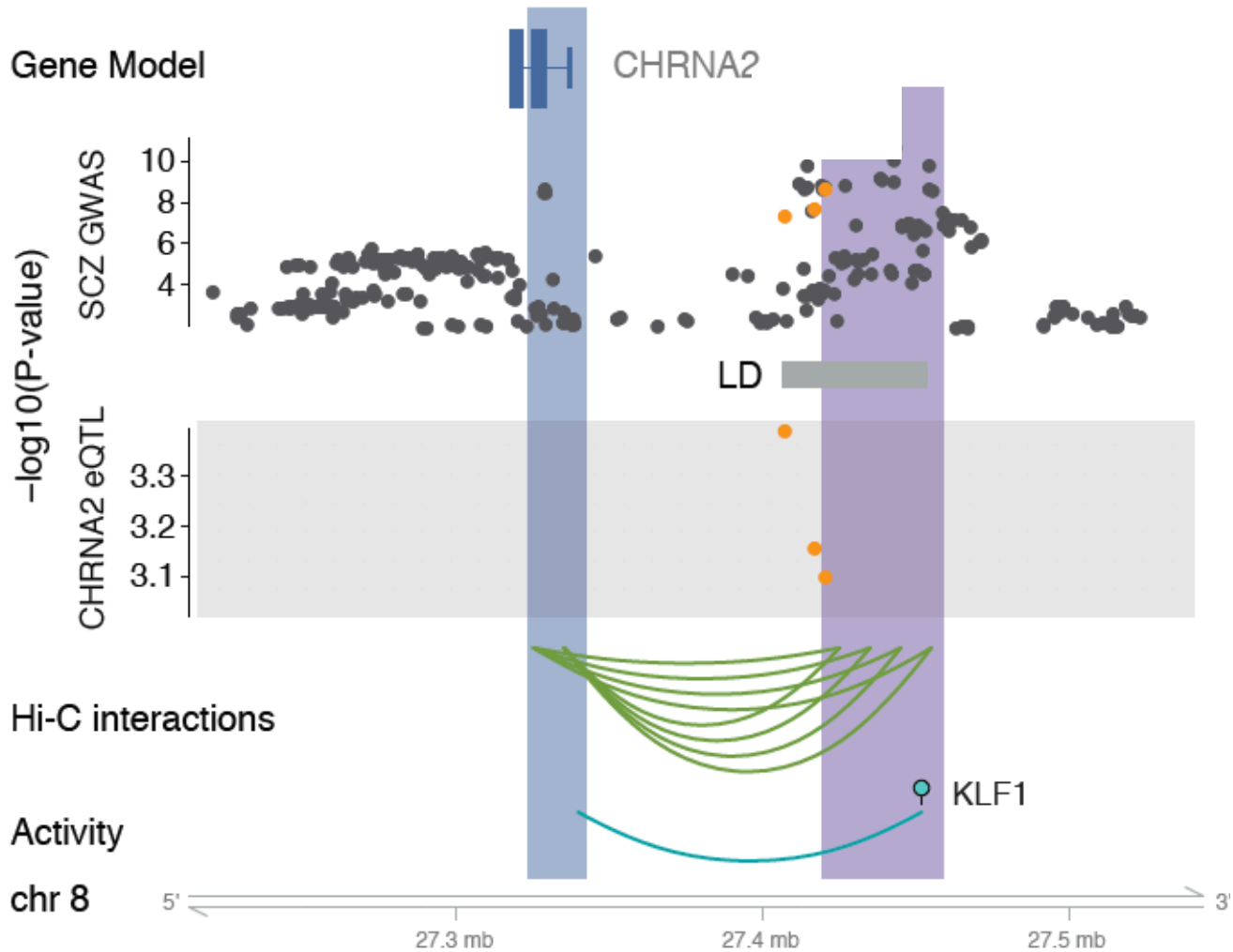
Gene regulatory network inference from Hi-C, QTLs & Activity Correlations

Imputed gene regulatory network for the human brain



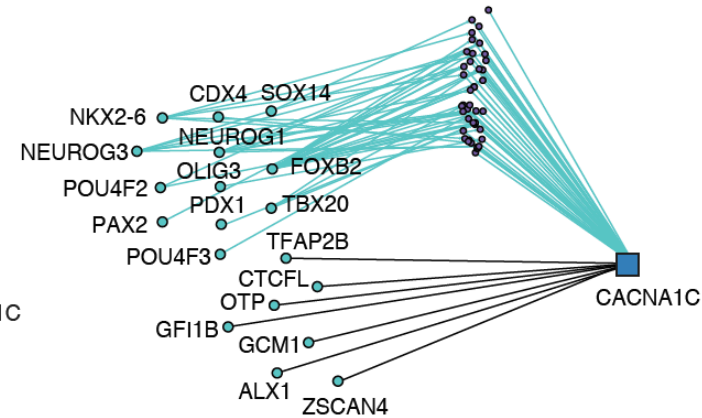
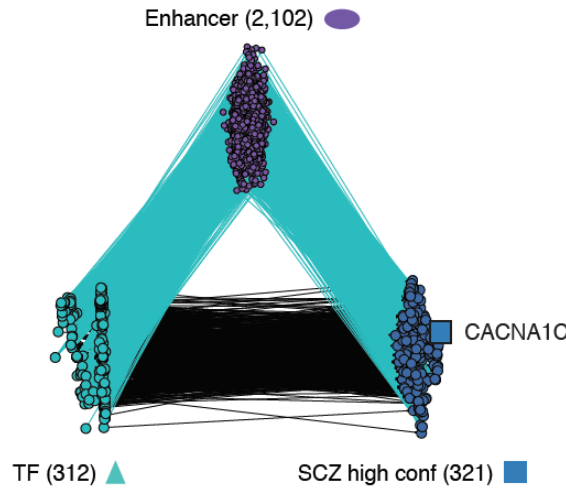
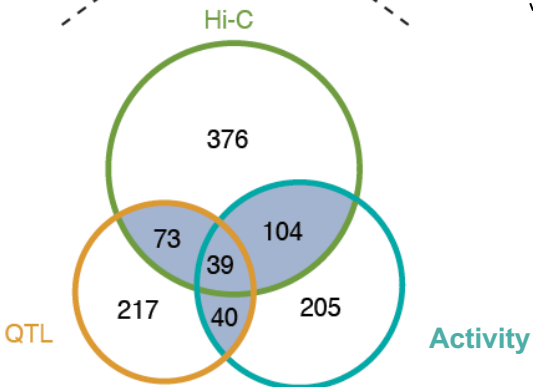
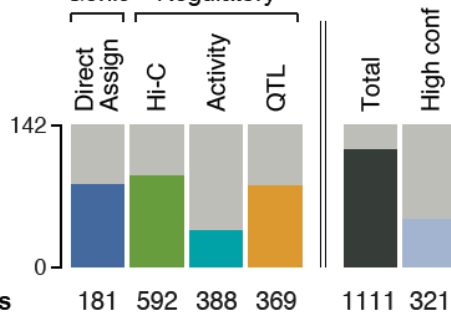
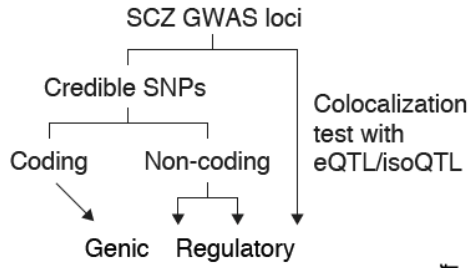
subnetworks targeting single cell marker genes

GWAS variants for SCZ genes



142

Linking GWAS SNPs to disease genes using the regulatory network



321
high-confident
SCZ genes

Interpretable deep learning model, embedding the gene regulatory network, for understanding functional genomics in neuropsychiatric disorders

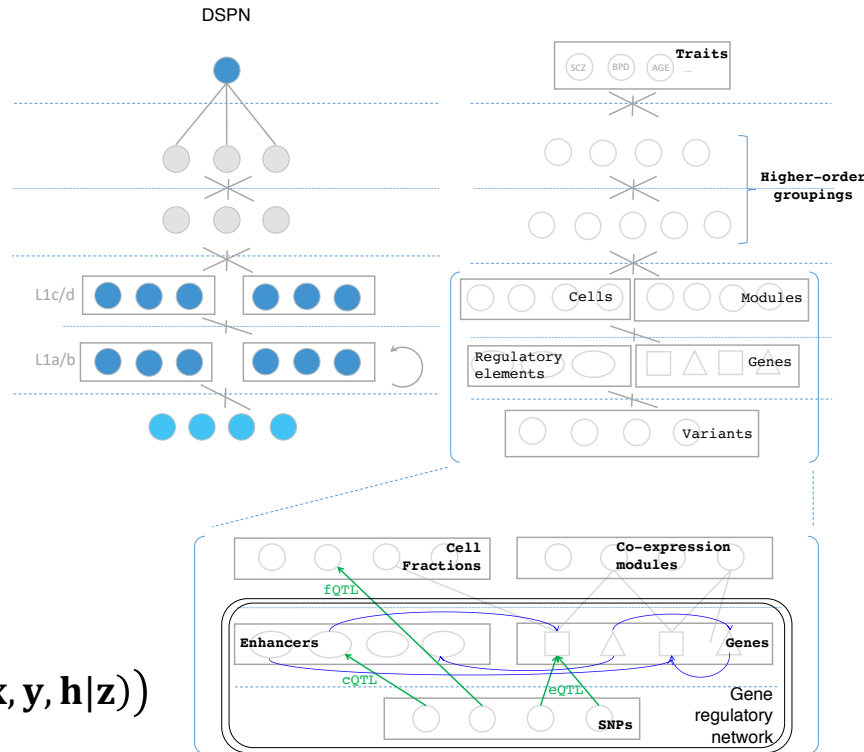
- Construction of an adult brain resource with 1866 individuals
- Using the changing proportions of cell types (via **single-cell deconvolution**) to account for expression variation across a population
- Large-scale processing defines ~79K PFC **enhancers & creates a comprehensive QTL** resource (~2.5M eQTLs + cQTLs & fQTLs)
- Connecting QTLs, enhancer activity relationships & Hi-C contacts into a **brain regulatory network** & using this to link SCZ GWAS SNPs to genes
- Embedding the reg. network in a **deep-learning model** to predict psychiatric disease from genotype & transcriptome. Using this to suggest specific pathways & genes, as potential drug targets.
- Other resource uses: highlighting aging related genes

Deep Structured Phenotype Network (DSPN)

Gene regulatory network builds skeleton

Energy model:

$$p(\mathbf{x}, \mathbf{y}, \mathbf{h} | \mathbf{z}) \propto \exp(-E(\mathbf{x}, \mathbf{y}, \mathbf{h} | \mathbf{z}))$$



Boltzmann machine

y: phenotypes

h: hidden units (e.g., circuits)

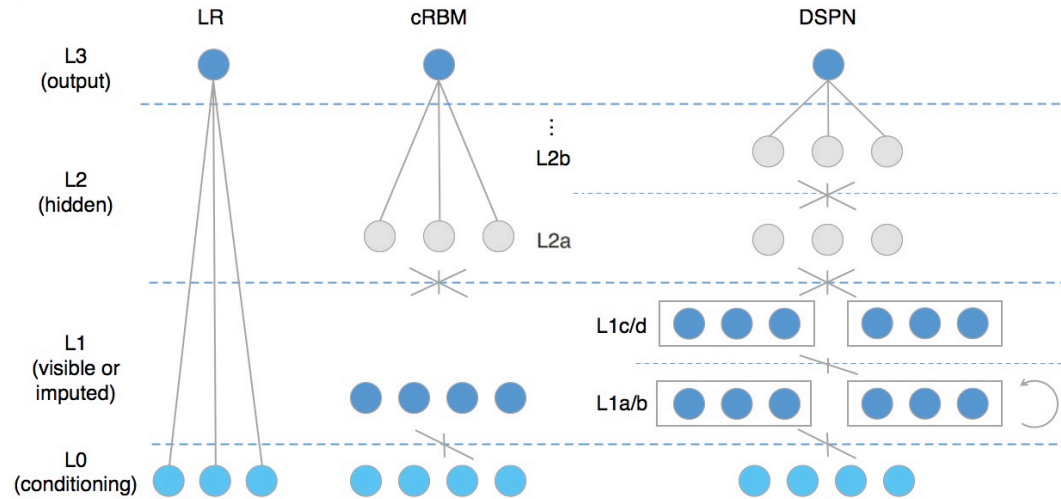
x: intermediate phenotypes (e.g., genes, enhancers)

z: genotypes (e.g., SNPs)

W: weights (e.g., regulatory network)

$$E(\mathbf{x}, \mathbf{y}, \mathbf{h} | \mathbf{z}) = -\mathbf{z}^T \mathbf{W}_1 \mathbf{x} - \mathbf{x}^T \mathbf{W}_2 \mathbf{x} - \mathbf{x}^T \mathbf{W}_3 \mathbf{h} - \mathbf{h}^T \mathbf{W}_4 \mathbf{h} - \mathbf{h}^T \mathbf{W}_5 \mathbf{y} - \text{Bias}$$

DSPN improves brain disease prediction by adding deep layers



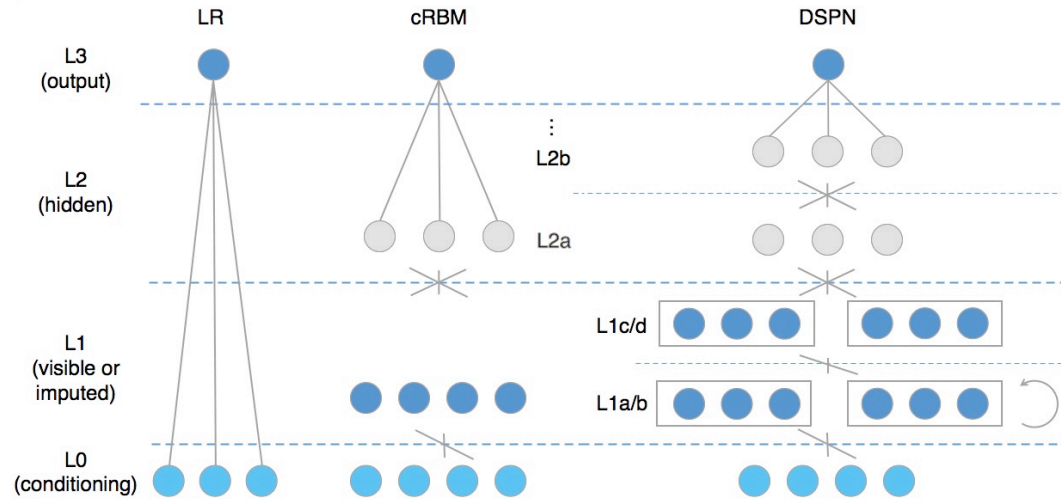
Method	LR-genotype	LR-transcriptome	cRBM	DSPN-imputation	DSPN-full
Schizophrenia	54.6%	63.0%	70.0%	59.0%	73.6%
Bipolar Disorder	56.7%	63.3%	71.1%	67.2%	76.7%
Autism Spectrum Disorder	50.0%	51.7%	67.2%	62.5%	68.3%



X 6.0

Accuracy = chance to correctly predict disease/health

DSPN improves brain disease prediction by adding deep layers



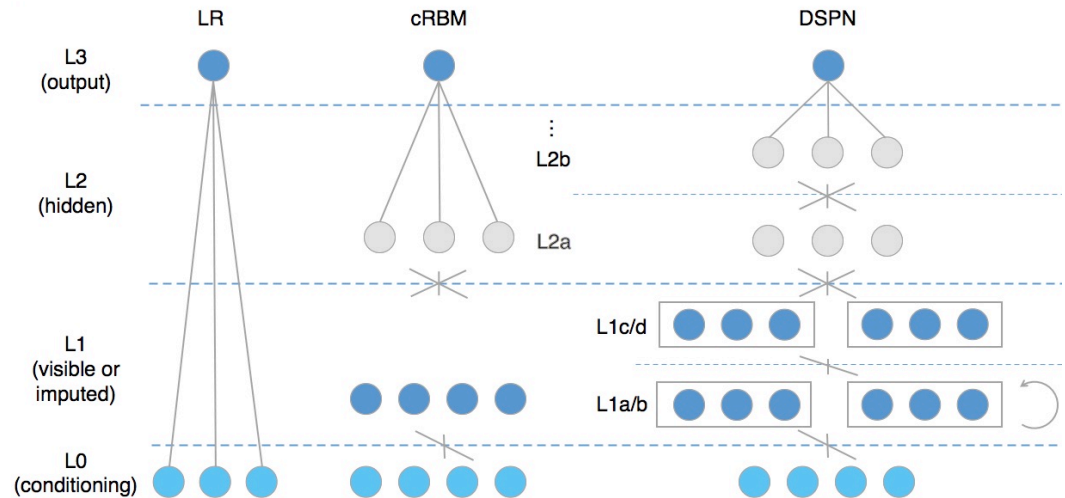
Method	LR-genotype	LR-transcriptome	cRBM	DSPN-imputation	DSPN-full
Schizophrenia	54.6%	63.0%	70.0%	59.0%	73.6%
Bipolar Disorder	56.7%	63.3%	71.1%	67.2%	76.7%
Autism Spectrum Disorder	50.0%	51.7%	67.2%	62.5%	68.3%



X 2.5

Accuracy = chance to correctly predict disease/health

DSPN improves brain disease prediction by adding deep layers



Method	LR-genotype	LR-transcriptome	cRBM	DSPN-imputation	DSPN-full
Schizophrenia	54.6%	63.0%	70.0%	59.0%	73.6%
Bipolar Disorder	56.7%	63.3%	71.1%	67.2%	76.7%
Autism Spectrum Disorder	50.0%	51.7%	67.2%	62.5%	68.3%

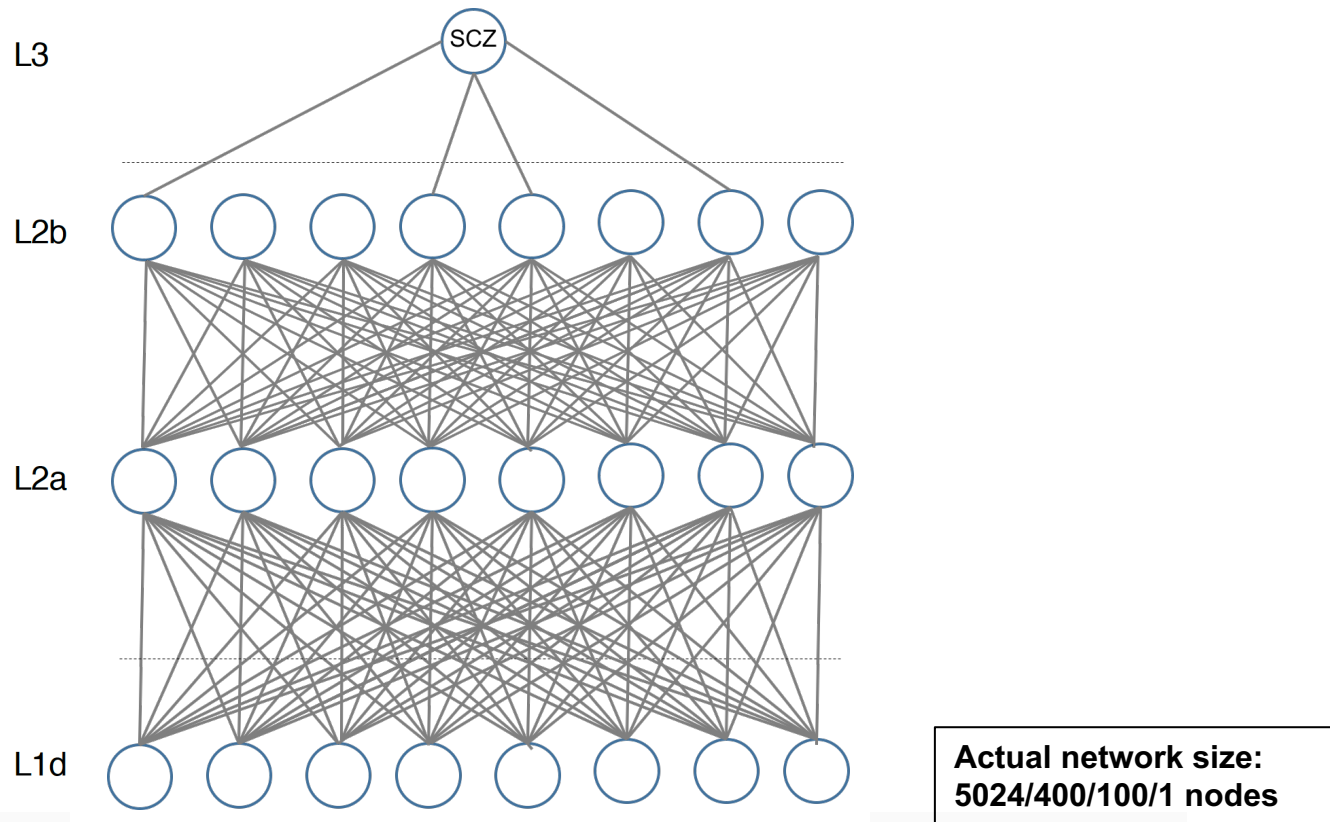
X 3.1

Accuracy = chance to correctly predict disease/health

Interpretable deep learning model, embedding the gene regulatory network, for understanding functional genomics in neuropsychiatric disorders

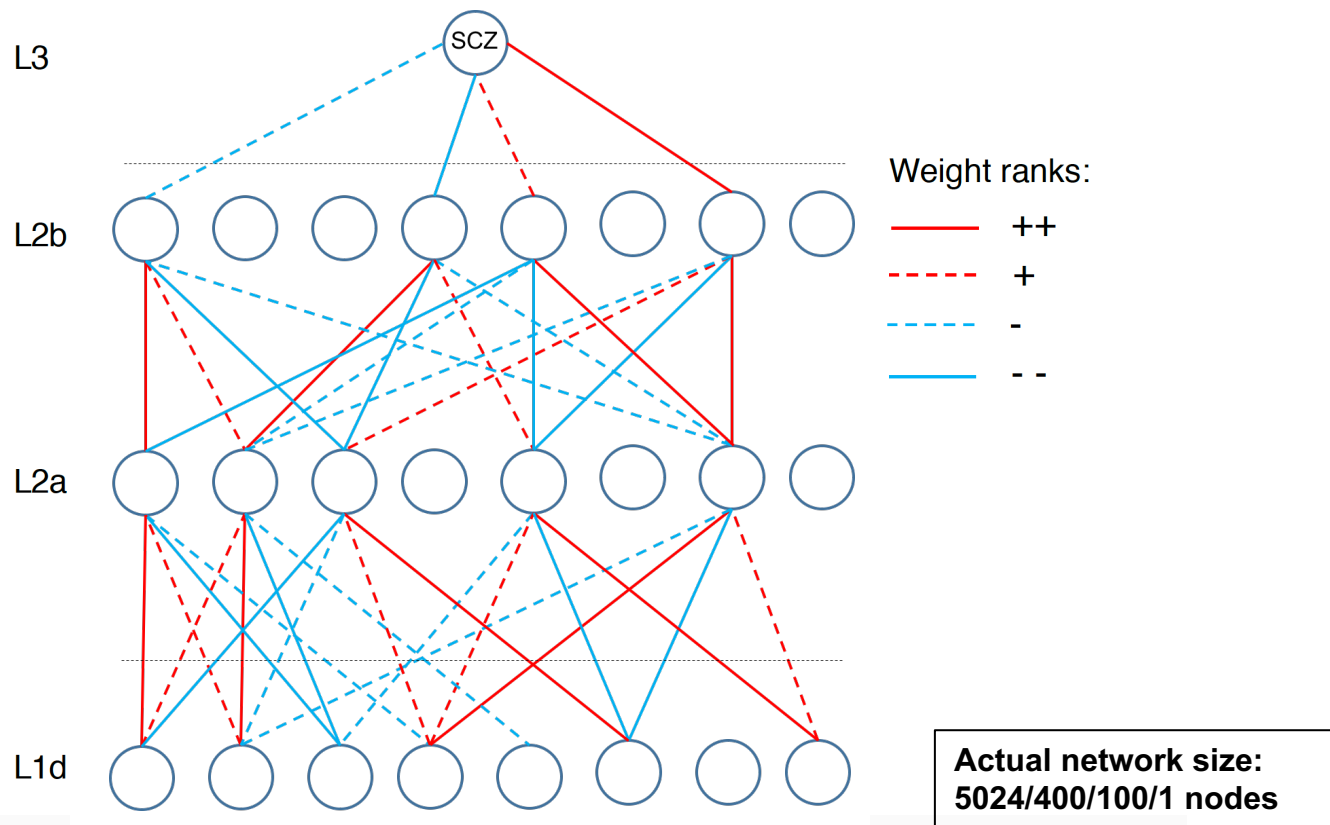
- Construction of an adult brain resource with 1866 individuals
- Using the changing proportions of cell types (via **single-cell deconvolution**) to account for expression variation across a population
- Large-scale processing defines ~79K PFC **enhancers & creates a comprehensive QTL** resource (~2.5M eQTLs + cQTLs & fQTLs)
- Connecting QTLs, enhancer activity relationships & Hi-C contacts into a **brain regulatory network** & using this to link SCZ GWAS SNPs to genes
- Embedding the reg. network in a **deep-learning model** to predict psychiatric disease from genotype & transcriptome. Using this to suggest specific pathways & genes, as potential drug targets.
- Other resource uses: highlighting aging related genes

Multilevel Network Interpretation



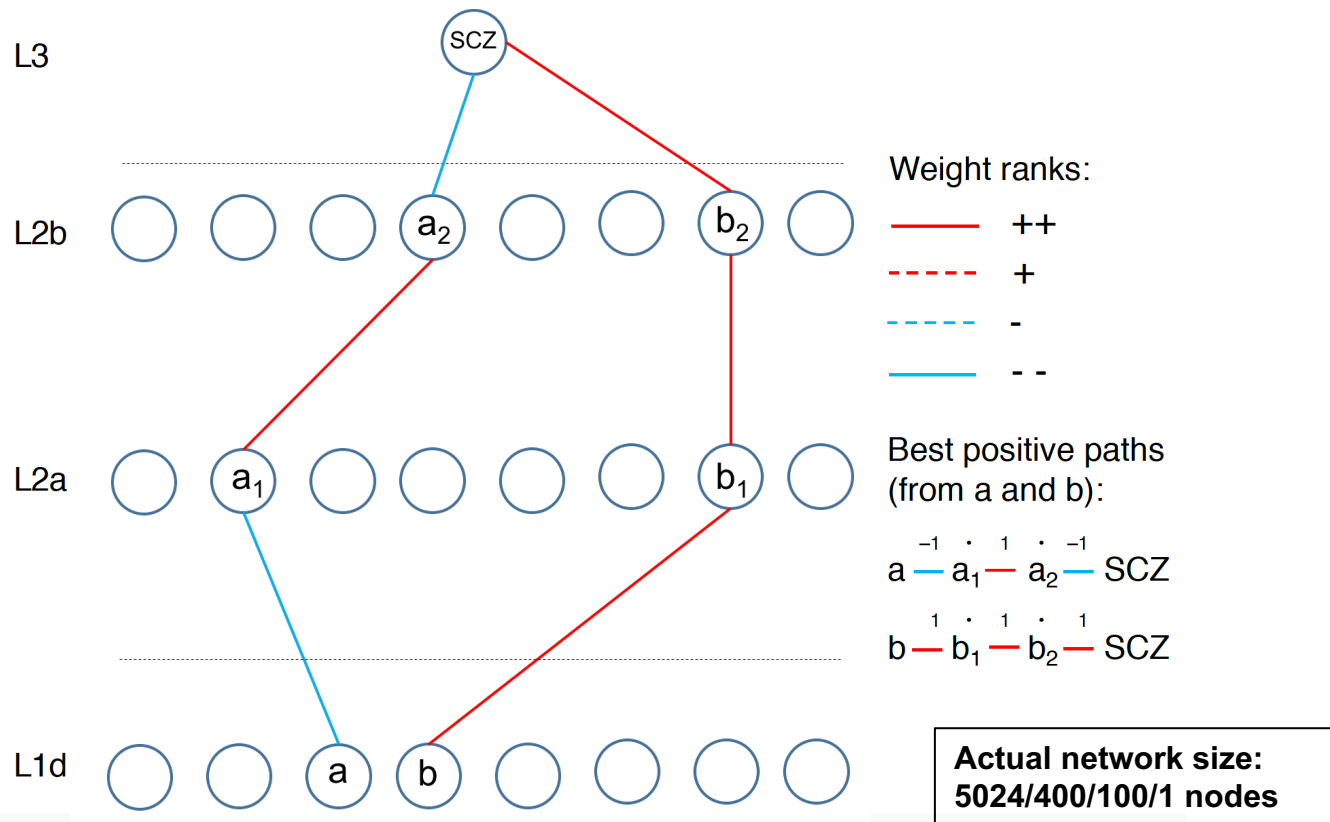
- Start with a fully connected trained network

Multilevel Network Interpretation



- Start with a fully connected trained network
- Sparsify network using edges with largest absolute weights (+/-)

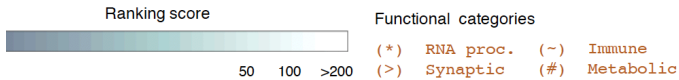
Multilevel Network Interpretation



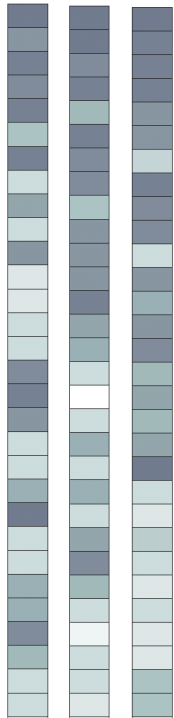
- Start with a fully connected trained network
- Sparsify network using edges with largest absolute weights (+/-)
- Extract 'best positive paths' to each prioritized module (e.g. a-a₁-a₂-SCZ) by summing weights and multiplying signs

DSPN discovers enriched pathways and linkages to genetic variation

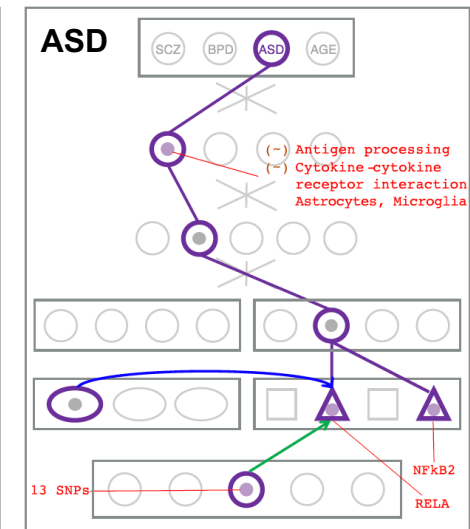
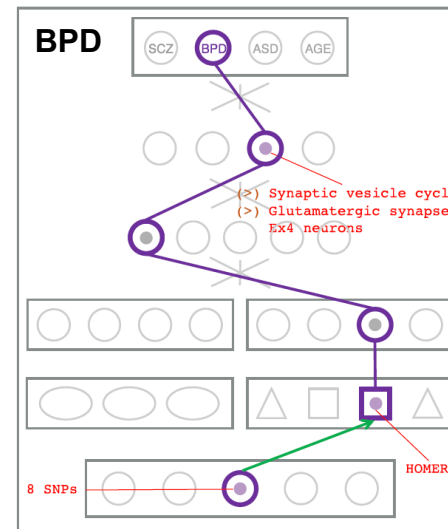
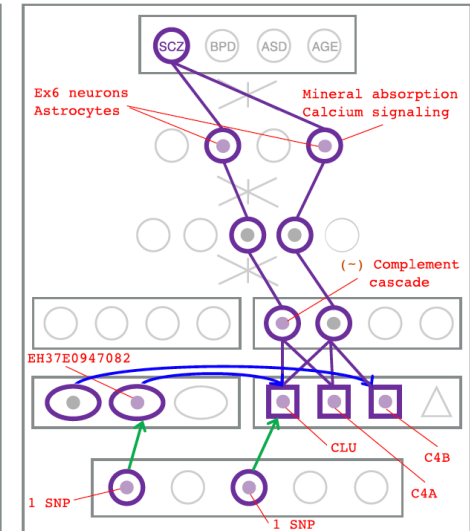
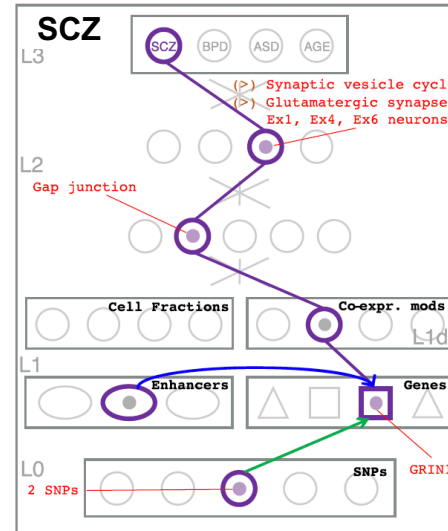
Cross-disorder MOD/HOG enrichment ranking



SCZ BPD ASD

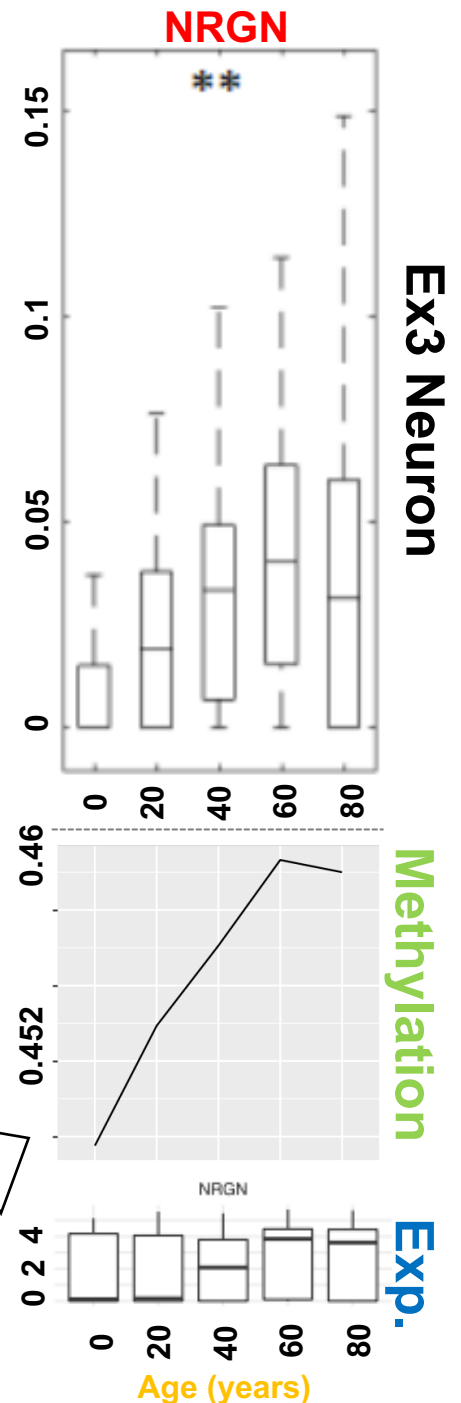
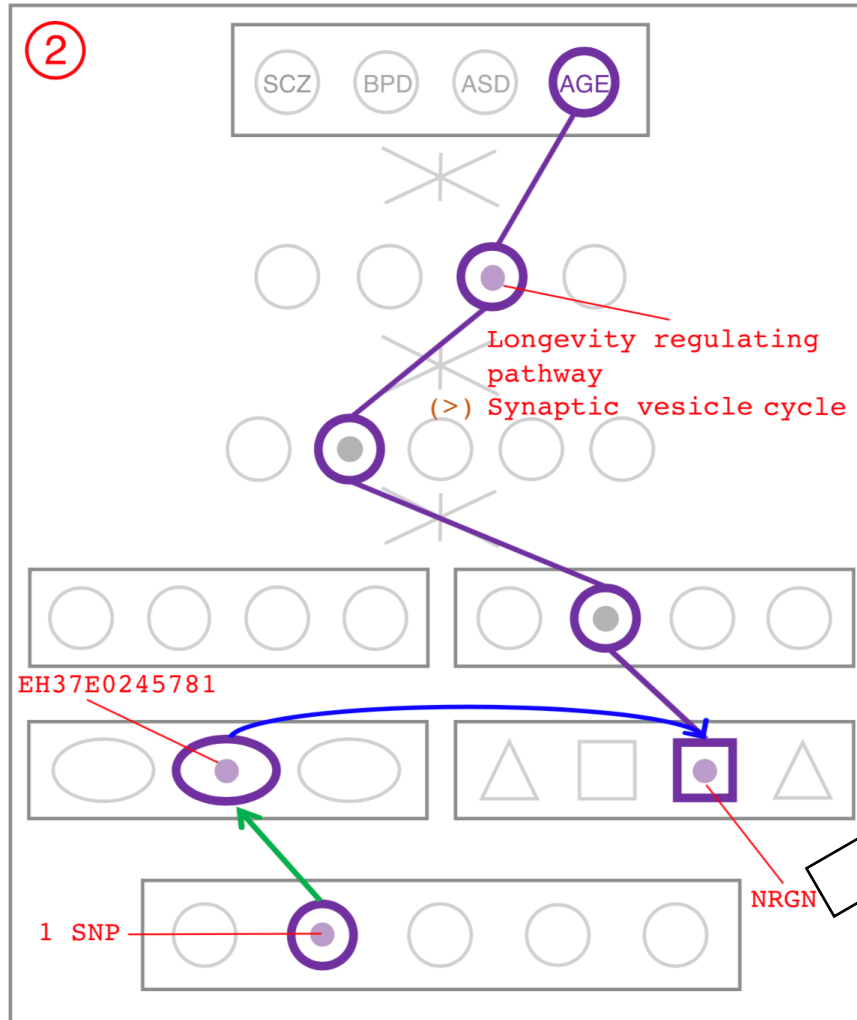


- (*) Spliceosome / RNA splicing
- (>) Synaptic vesicle cycle
- (~) Antigen proc. and presentation
- Vesicle localization
- Proteasome
- (*) mRNA processing
- Chromatin modification
- (#) Oxidative phosphorylation
- Retrograde endocannabinoid sig.
- (>) Chemical synaptic transmission
- Peptidyl-lysine modification
- Endocytosis
- Ubiquitin mediated proteolysis
- (>) Anterograde trans-synaptic sig.
- (*) mRNA transport
- Phosphatidylinositol signaling
- Hippo signaling pathway
- (~) Staph./ Epstein-Barr virus inf.
- (>) Synaptic signaling
- Autophagy
- (>) Dop./GABA/Glutamatergic synapse
- (>) Calcium signaling
- (>) Endocrine calcium reabsorption
- (*) RNA degradation / transport
- (#) Ribosome
- Neuron projection morphogenesis
- (~) Fc receptor signaling pathway
- cGMP-PKG signaling pathway
- (~) mTOR signaling pathway
- (~) Cytokine-cytokine receptor int.



NRGN has variable expression over age and is in Synaptic vesicle cycle pathway is enriched in SCZ, BPD, ASD

NRGN is a gene associated with the **Synaptic vesicle pathway** and **NRGN expression** and **methylation** is correlated with **Age**



Interpretable deep learning model, embedding the gene regulatory network, for understanding functional genomics in neuropsychiatric disorders

- Construction of an adult brain resource with 1866 individuals
- Using the changing proportions of cell types (via **single-cell deconvolution**) to account for expression variation across a population
- Large-scale processing defines ~79K PFC **enhancers & creates a comprehensive QTL** resource (~2.5M eQTLs + cQTLs & fQTLs)
- Connecting QTLs, enhancer activity relationships & Hi-C contacts into a **brain regulatory network** & using this to link SCZ GWAS SNPs to genes
- Embedding the reg. network in a **deep-learning model** to predict psychiatric disease from genotype & transcriptome. Using this to suggest specific pathways & genes, as potential drug targets.
- Other resource uses: highlighting aging related genes

Interpretable deep learning model, embedding the gene regulatory network, for understanding functional genomics in neuropsychiatric disorders

- Construction of an adult brain resource with 1866 individuals
- Using the changing proportions of cell types (via **single-cell deconvolution**) to account for expression variation across a population
- Large-scale processing defines ~79K PFC **enhancers & creates a comprehensive QTL** resource (~2.5M eQTLs + cQTLs & fQTLs)
- Connecting QTLs, enhancer activity relationships & Hi-C contacts into a **brain regulatory network** & using this to link SCZ GWAS SNPs to genes
- Embedding the reg. network in a **deep-learning model** to predict psychiatric disease from genotype & transcriptome. Using this to suggest specific pathways & genes, as potential drug targets.
- Other resource uses: highlighting aging related genes

Acknowledgments

Dedicated to
Pamela Sklar

NIMH:
Geetha Senthil
Lora Bingaman
David Panchision
Alexander Arguello
Thomas Lehner

Daifeng Wang, Shuang Liu, Jonathan Warrell, Hyejung Won, Xu Shi, Fabio Navarro, Declan Clarke, Mengting Gu,

Prashant Emani, Yucheng T. Yang, Min Xu, Michael Gandal, Shaoke Lou, Jing Zhang, Jonathan J. Park, Chengfei Yan, Suhn Kyong Rhie, Kasidet Manakongtreecheep, Holly Zhou, Aparna Nathan, Mette Peters, Eugenio Mattei, Dominic Fitzgerald, Tonya Brunetti, Jill Moore, Yan Jiang, Kiran Girdhar, Gabriel Hoffman, Selim Kalayci, Zeynep Hulya Gumus, Greg Crawford,

PsychENCODE Consortium,

Panos Roussos, Schahram Akbarian, Andrew E. Jaffe, Kevin White, Zhiping Weng, Nenad Sestan,

Daniel H. Geschwind, James A. Knowles

The PsychENCODE Consortium: Allison E Ashley-Koch, Duke University; Gregory E Crawford, Duke University; Melanie E Garrett, Duke University; Lingyun Song, Duke University; Alexias Safi, Duke University; Graham D Johnson, Duke University; Gregory A Wray, Duke University; Timothy E Reddy, Duke University; Fernando S Goes, Johns Hopkins University; Peter Zandi, Johns Hopkins University; Julien Bryois, Karolinska Institutet; Andrew E Jaffe, Lieber Institute for Brain Development; Amanda J Price, Lieber Institute for Brain Development; Nikolay A Ivanov, Lieber Institute for Brain Development; Leonardo Collado-Torres, Lieber Institute for Brain Development; Thomas M Hyde, Lieber Institute for Brain Development; Emily E Burke, Lieber Institute for Brain Development; Joel E Kleiman, Lieber Institute for Brain Development; Ran Tao, Lieber Institute for Brain Development; Joo Heon Shin, Lieber Institute for Brain Development; Schahram Akbarian, Icahn School of Medicine at Mount Sinai; Kiran Girdhar, Icahn School of Medicine at Mount Sinai; Yan Jiang, Icahn School of Medicine at Mount Sinai; Marija Kundakovic, Icahn School of Medicine at Mount Sinai; Leanne Brown, Icahn School of Medicine at Mount Sinai; Bibi S Kassim, Icahn School of Medicine at Mount Sinai; Royce B Park, Icahn School of Medicine at Mount Sinai; Jennifer R Wiseman, Icahn School of Medicine at Mount Sinai; Elizabeth Zharovsky, Icahn School of Medicine at Mount Sinai; Rivka Jacobov, Icahn School of Medicine at Mount Sinai; Olivia Devillers, Icahn School of Medicine at Mount Sinai; Elie Flatow, Icahn School of Medicine at Mount Sinai; Gabriel E Hoffman, Icahn School of Medicine at Mount Sinai; Barbara K Lipska, Human Brain Collection Core, National Institutes of Health, Bethesda, MD; David A Lewis, University of Pittsburgh; Vahram Haroutunian, Icahn School of Medicine at Mount Sinai and James J Peters VA Medical Center; Chang-Gyu Hahn, University of Pennsylvania; Alexander W Charney, Mount Sinai; Stella Dracheva, Mount Sinai; Alexey Kozenkov, Mount Sinai; Judson Belmont, Icahn School of Medicine at Mount Sinai; Diane DelValle, Icahn School of Medicine at Mount Sinai; Nancy Francoeur, Icahn School of Medicine at Mount Sinai; Evi Hadjimichael, Icahn School of Medicine at Mount Sinai; Dalila Pinto, Icahn School of Medicine at Mount Sinai; Harm van Bakel, Icahn School of Medicine at Mount Sinai; Panos Roussos, Mount Sinai; John F Fullard, Mount Sinai; Jaroslav Bendl, Mount Sinai; Mads E Hauberg, Mount Sinai; Lara M Mangravite, Sage Bionetworks; Mette A Peters, Sage Bionetworks; Yooree Chae, Sage Bionetworks; Junmin Peng, St. Jude Children's Hospital; Mingming Niu, St. Jude Children's Hospital; Xusheng Wang, St. Jude Children's Hospital; Maree J Webster, Stanley Medical Research Institute; Thomas G Beach, Banner Sun Health Research Institute; Chao Chen, Central South University; Yi Jiang, Central South University; Rujia Dai, Central South University; Annie W Shieh, SUNY Upstate Medical University; Chunyu Liu, SUNY Upstate Medical University; Kay S Grennan, SUNY Upstate Medical University; Yan Xia, SUNY Upstate Medical University/Central South University; Ramu Vadukapuram, SUNY Upstate Medical University; Yongjun Wang, Central South University; Dominic Fitzgerald, The University of Chicago; Lijun Cheng, The University of Chicago; Miguel Brown, The University of Chicago; Mimi Brown, The University of Chicago; Tonya Brunetti, The University of Chicago; Thomas Goodman, The University of Chicago; Majd Alsayed, The University of Chicago; Michael J Gandal, University of California, Los Angeles; Daniel H Geschwind, University of California, Los Angeles; Hyejung Won, University of California, Los Angeles; Damon Polioudakis, University of California, Los Angeles; Brie Wamsley, University of California, Los Angeles; Jiani Yin, University of California, Los Angeles; Tarik Hadzic, University of California, Los Angeles; Luis De La Torre Ubieta, UCLA; Vivek Swarup, University of California, Los Angeles; Stephan J Sanders, University of California, San Francisco; Matthew W State, University of California, San Francisco; Donna M Werling, University of California, San Francisco; Joon-Yong An, University of California, San Francisco; Brooke Sheppard, University of California, San Francisco; A Jeremy Willsey, University of California, San Francisco; Kevin P White, The University of Chicago; Mohana Ray, The University of Chicago; Gina Giase, SUNY Upstate Medical University; Amira Kefi, University of Illinois at Chicago; Eugenio Mattei, University of Massachusetts Medical School; Michael Purcaro, University of Massachusetts Medical School; Zhiping Weng, University of Massachusetts Medical School; Jill Moore, University of Massachusetts Medical School; Henry Pratt, University of Massachusetts Medical School; Jack Huey, University of Massachusetts Medical School; Tyler Borrmann, University of Massachusetts Medical School; Patrick F Sullivan, University of North Carolina - Chapel Hill; Paola Giusti-Rodriguez, University of North Carolina - Chapel Hill; Yunjung Kim, University of North Carolina - Chapel Hill; Patrick Sullivan, University of North Carolina - Chapel Hill; Jin Szatkiewicz, University of North Carolina - Chapel Hill; Suhn Kyong Rhie, University of Southern California; Christopher Armoskus, University of Southern California; Adrian Camarena, University of Southern California; Peggy J Farnham, University of Southern California; Valeria N Spitsyna, University of Southern California; Heather Witt, University of Southern California; Shannon Schreiner, University of Southern California; Oleg V Evgrafov, SUNY Downstate Medical Center; James A Knowles, SUNY Downstate Medical Center; Mark Gerstein, Yale University; Shuang Liu, Yale University; Daifeng Wang, Stony Brook University; Fabio C. P. Navarro, Yale University; Jonathan Warrell, Yale University; Declan Clarke, Yale University; Prashant S. Emani, Yale University; Mengting Gu, Yale University; Xu Shi, Yale University; Min Xu, Yale University; Yucheng T. Yang, Yale University; Robert R. Kitchen, Yale University; Gamze Gürsoy, Yale University; Jing Zhang, Yale University; Becky C Carlyle, Yale University; Angus C Nairn, Yale University; Sirisha Pochareddy, Yale University; Nenad Sestan, Yale University; Mario Skarica, Yale University; Zhen Li, Yale University; Andre M.M. Sousa, Yale University; Gabriel Santpere, Yale University; Jinmyung Choi, Yale University; Ying Zhu, Yale University; Tianliuyun Gao, Yale University; Daniel J Miller, Yale University; Adriana Cherskov, Yale University; Mo Yang, Yale University; Anahita Amiri, Yale University; Gianfilippo Coppola, Yale University; Jessica Mariani, Yale University; Soraya Scuderi, Yale University; Anna Szekely, Yale University; Flora M Vaccarino, Yale University; Feinan Wu, Yale University; Sherman Weissman, Yale University; Tanmoy Roychowdhury, Mayo Clinic Rochester; Alexej Abyzov, Mayo Clinic Rochester.

Extra



Info about content in this slide pack

- General PERMISSIONS
 - This Presentation is copyright Mark Gerstein, Yale University, 2019.
 - Please read permissions statement at www.gersteinlab.org/misc/permissions.html .
 - Feel free to use slides & images in the talk with PROPER acknowledgement (via citation to relevant papers or link to gersteinlab.org).
 - Paper references in the talk were mostly from Papers.GersteinLab.org.
- PHOTOS & IMAGES. For thoughts on the source and permissions of many of the photos and clipped images in this presentation see <http://streams.gerstein.info> .
 - In particular, many of the images have particular EXIF tags, such as kwpotppt , that can be easily queried from flickr, viz: <http://www.flickr.com/photos/mbgmbg/tags/kwpotppt>