

Motivation

- **Creating a Brain-Specific Genomics Resource**
 - Previous development of Generic Resources to annotate the whole human genome (ENCODE, Roadmap, GTEx, 1000 Genomes, &c);
Not focused on the brain
 - Desire to provide a specific resource for the brain
 - Brain enhancers, specific splice forms, regulatory networks, & eQTLs
- Questions posed by this resource
 - **Are gene regulation, transcription and other aspects of genome function similar between the brain and other tissues (e.g., the liver)?**
 - **How do genomic variants affect brain phenotypes?
Can we make predictive models to explain this?**

Capstone #4 – talk outline

- Resource
 - Brain enhancers, QTLs, expression modules, networks &c
- Transcriptome analysis
 - 1) Spectral analysis of gene expression, unique brain expression patterns
 - 2) Non-coding transcription & retro-element activity, not atypical
- Integrative network models
 - Relating perturbations in gene expression to variants

Datasets

PsychENCODE

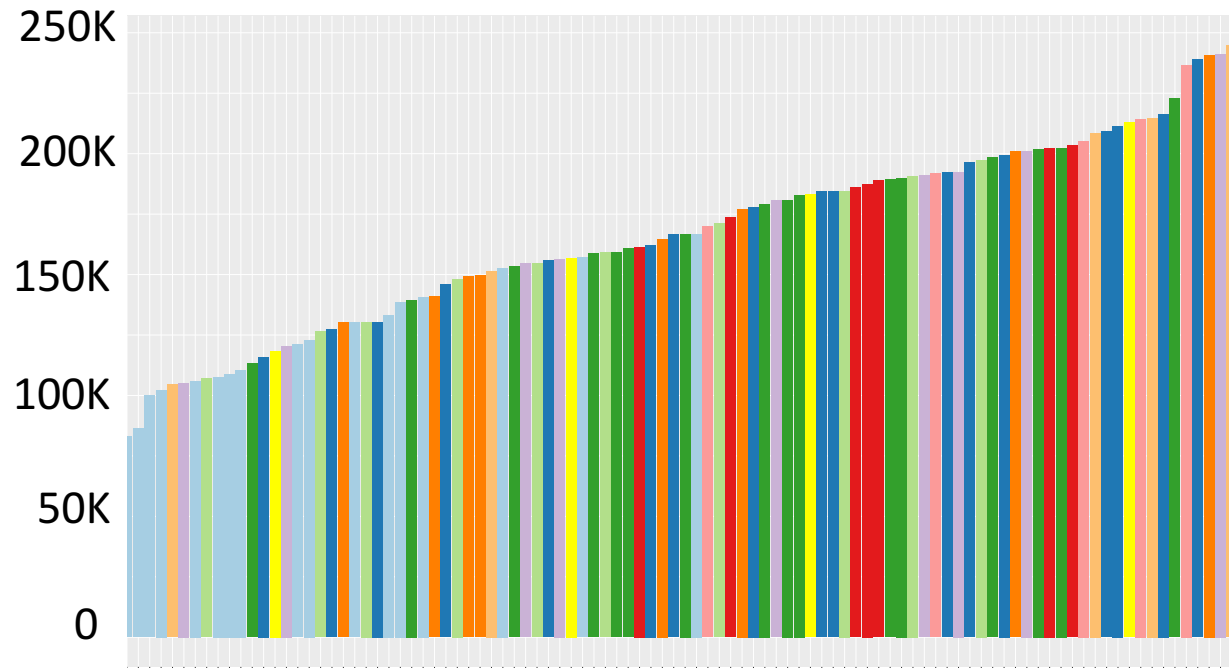
Datasets	Disease status	# samples	Brain region	Data Type
ENCODE/Roadmap	CTL		Multiple regions	ChIP-seq, RNA-seq
BrainSpan	CTL	606	16 regions	ChIP-seq, RNA-seq, genotype
CommonMind	SCZ, BD, CTL	613	DFC	RNA-seq, genotype
GTEEx	CTL	92	DFC	RNA-seq, genotype
BrainGVEX	SCZ, BD, CTL	429	DFC	RNA-seq, genotype
UCLA-ASD	ASD, CTL	253	DFC, CBC	ChIP-seq, RNA-seq, genotype
Yale-ASD	ASD, CTL	45	DFC, TC, V1, CBL	ChIP-seq, RNA-seq, genotype
CNON	SCZ, CTL	63	Olfactory Neuroepithelium	ChIP-seq
EpiGABA	CTL	4	OFC (BA11)	RNA-seq, genotype, ERRBS, Methylation
iPSC	CTL	51	iPSC	ChIP-seq, RNA-seq, WGS
TOTAL		~2156		

Brain Specific Enhancers and Co-expression Modules

Identifying brain-specific enhancers

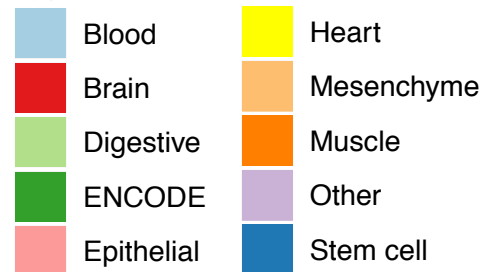
Identifying brain co-expression modules

of enhancers:

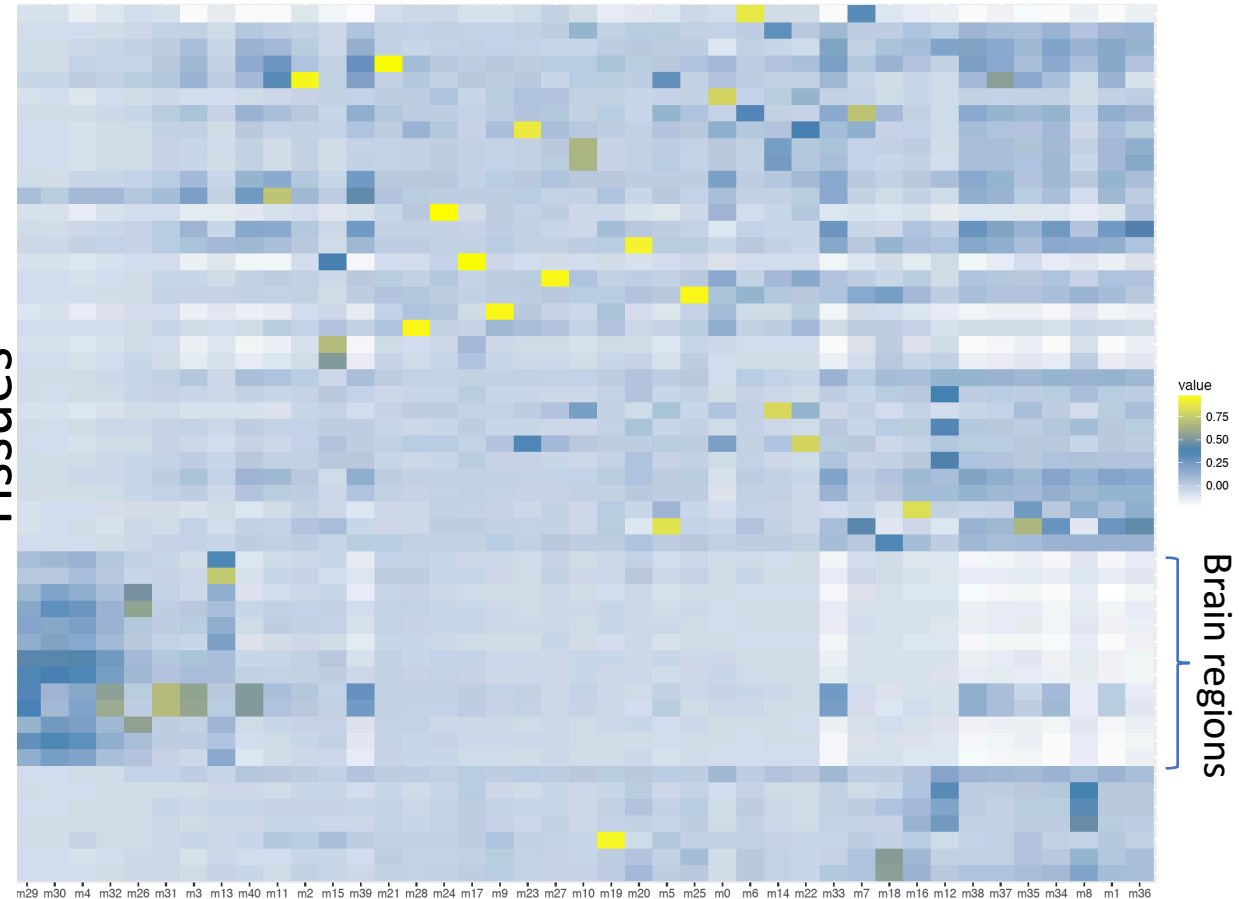


Tissues

Type

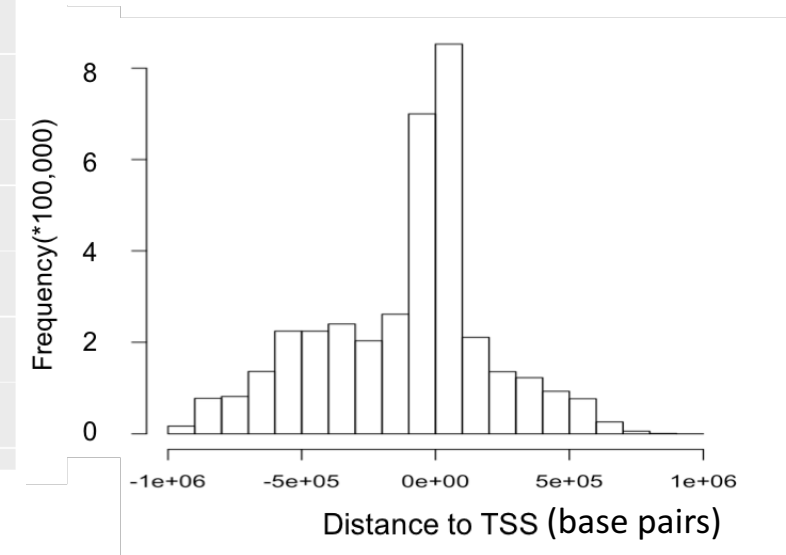
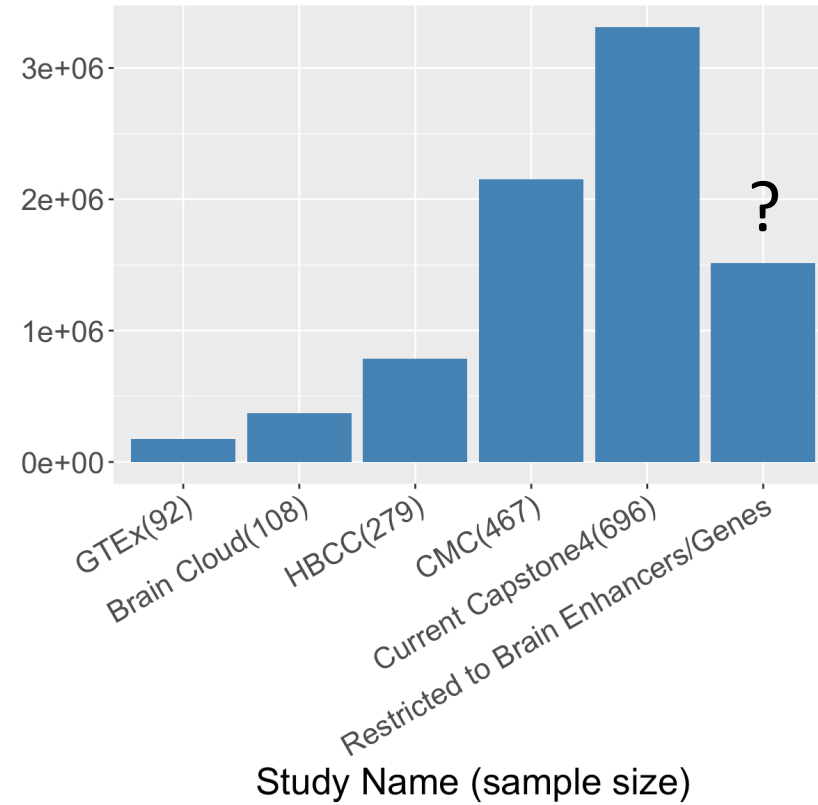
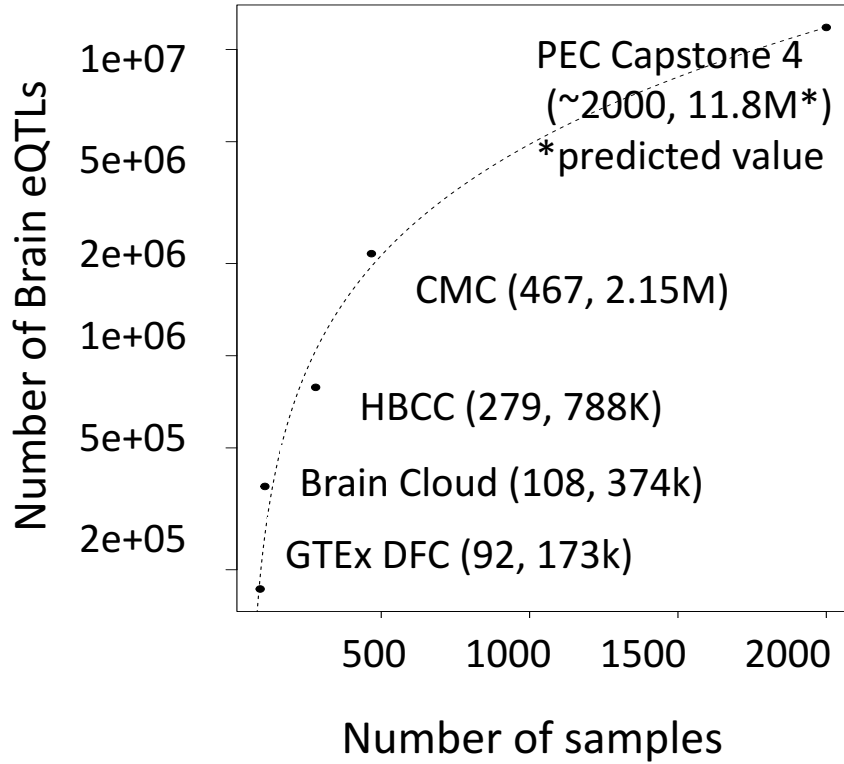


+/-ve corr:
p<0.001:



Modules (ordered by brain-specificity)

Brain QTLS



(Preliminary!) Characteristic Statistics for the Brain

	Total	Active in Brain regions	Brain Specific	Other tissue active (mean)	Other tissue specific (mean)
Enhancers ^[1]	2.3M	332.5K (14.5%)	130.2K (5.7%)	274.1K (11.9%)	83.5K (3.6%)
Genes ^[2]	23K	17.6K (76.4%)	386 (1.6%)	16.9K (73.5%)	25.9 (0.1%)
lncRNAs and pseudogenes ^[2]	33.2K	14K (42.2%)	814 (2.4%)	10K (30.1%)	136.7 (0.4%)
cis-eQTLs ^[2]	26.4M	3.25M (12.3%)	1.02M (3.9%)	599K (2.3%)	189K (0.73%)
Gene expression modules ^[2]	117	59 (50.4%)	6 (5.1%)	26.3 (22.4%)	3.3 (2.8%)
Differentially included exons ^[3]	2.5K	-	420 (16.4%)	-	142 (5.5%)

[1] Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature*, 2015.

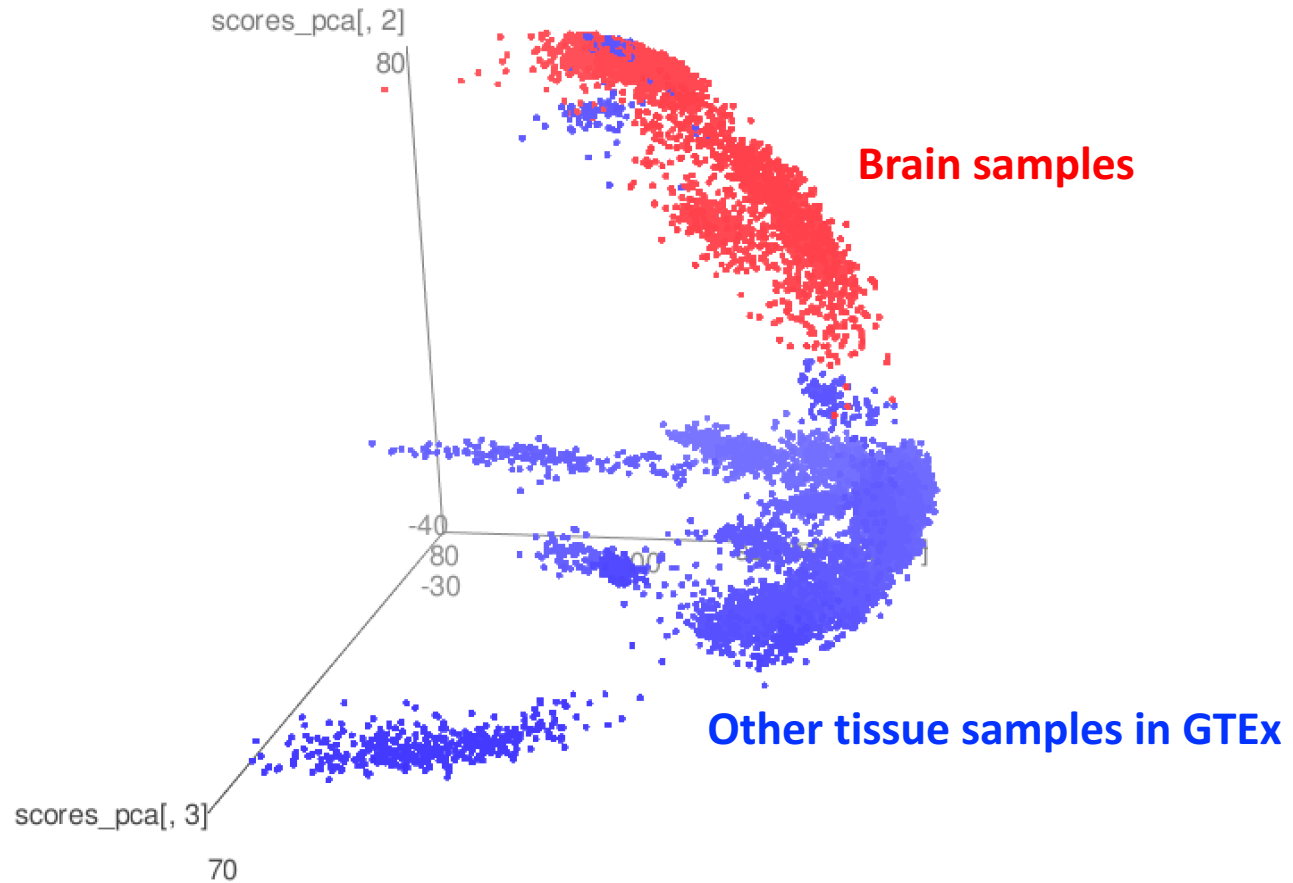
[2] The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, 2015.

[3] Melé *et al.* The human transcriptome across tissues and individuals. *Science*, 2015.

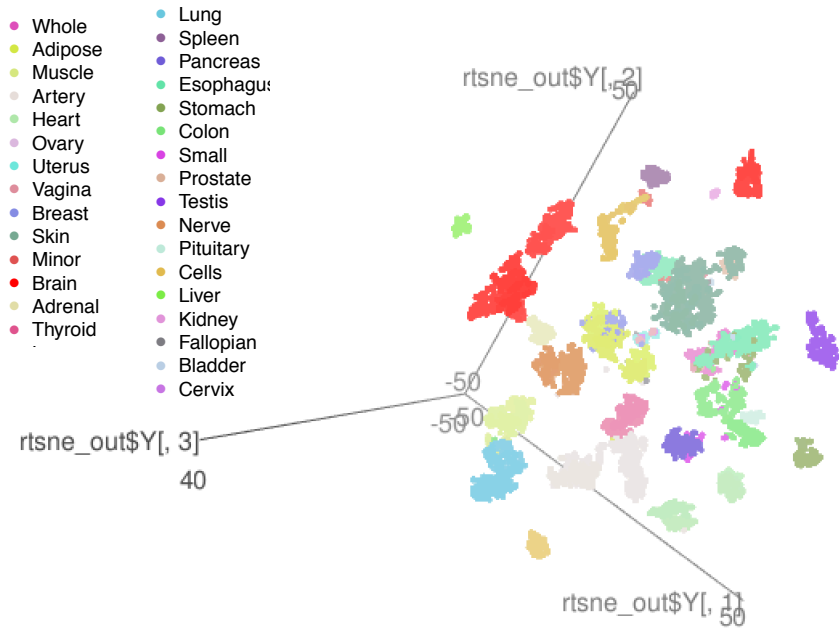
Capstone #4 – talk outline

- Resource
 - Brain enhancers, QTLs, expression modules, networks &c
- Transcriptome analysis
 - 1) Spectral analysis of gene expression, unique brain expression patterns
 - 2) Non-coding transcription & retro-element activity, not atypical
- Integrative network models
 - Relating perturbations in gene expression to variants

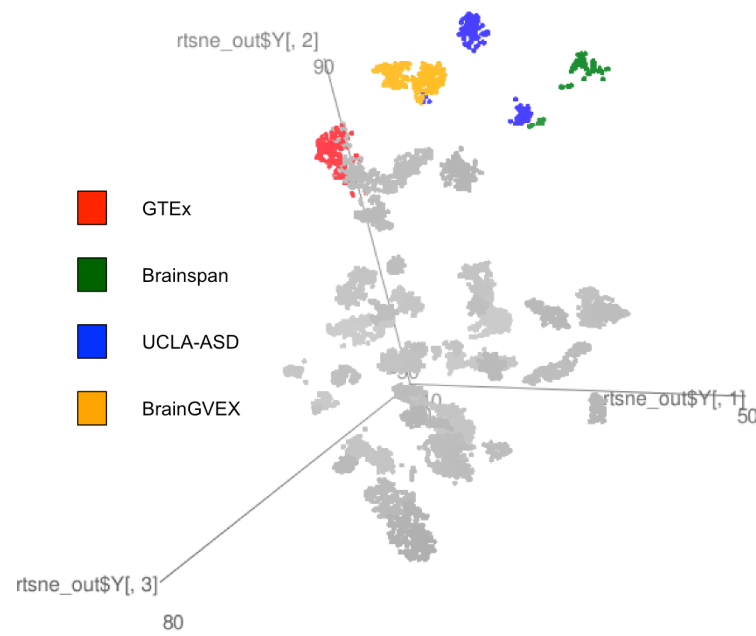
PCA of GTEX, GVEX, Brainspan, CMC, ASD



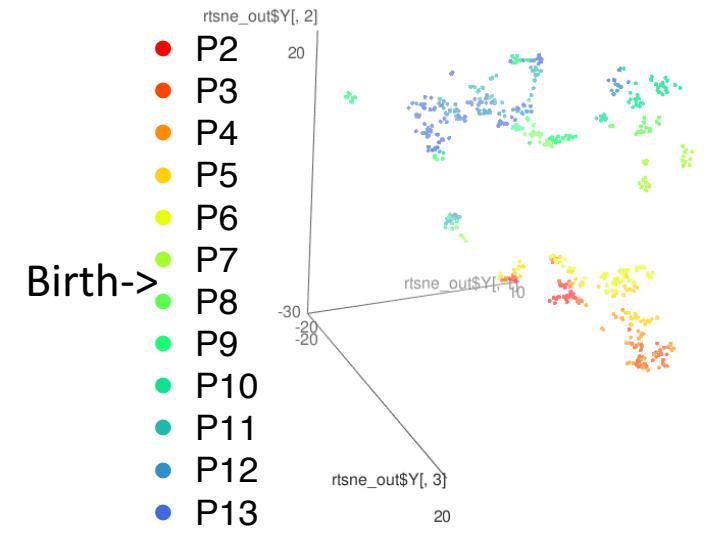
tSNE local clustering



GTEx

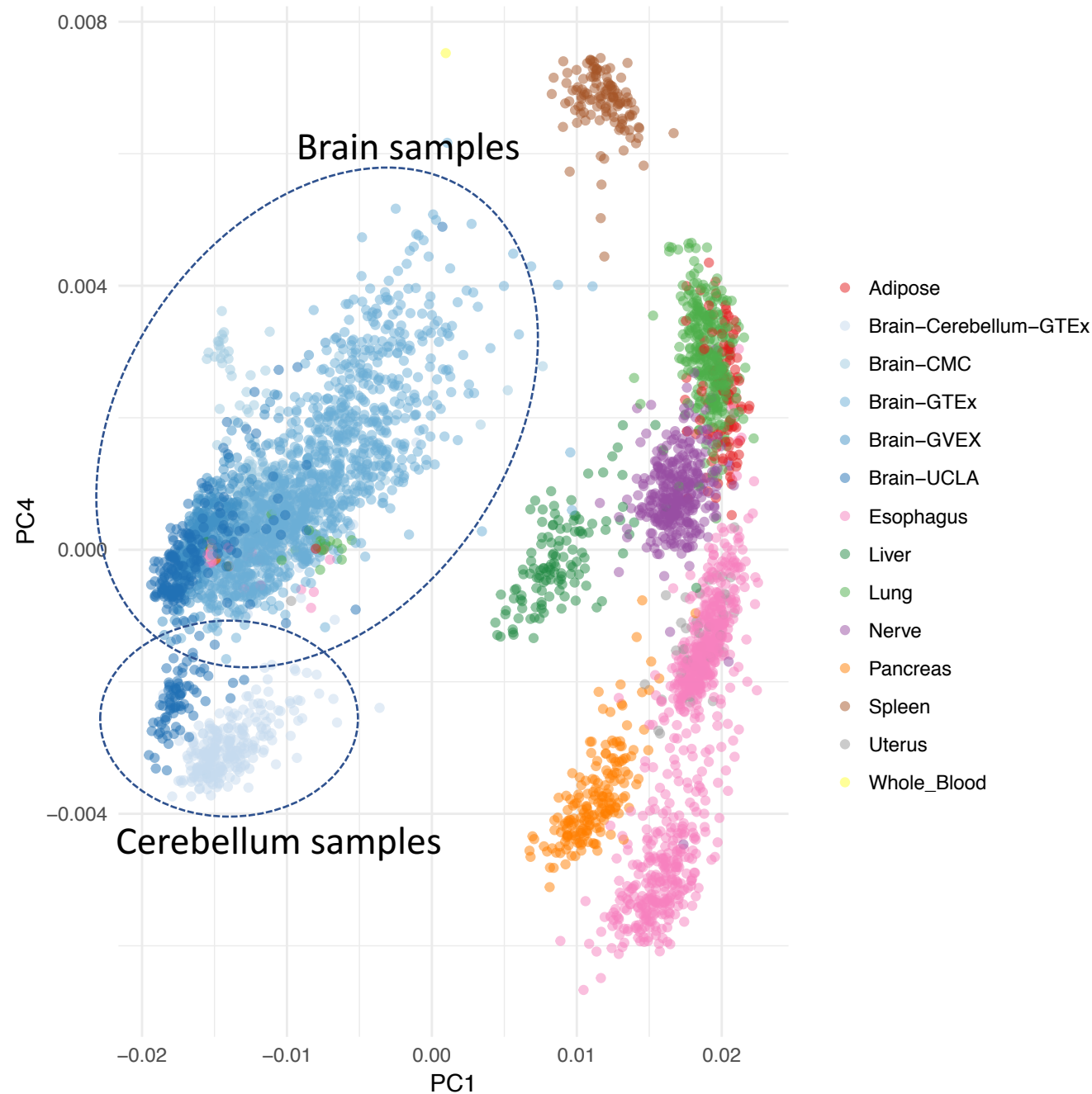


GTEx + some PEC

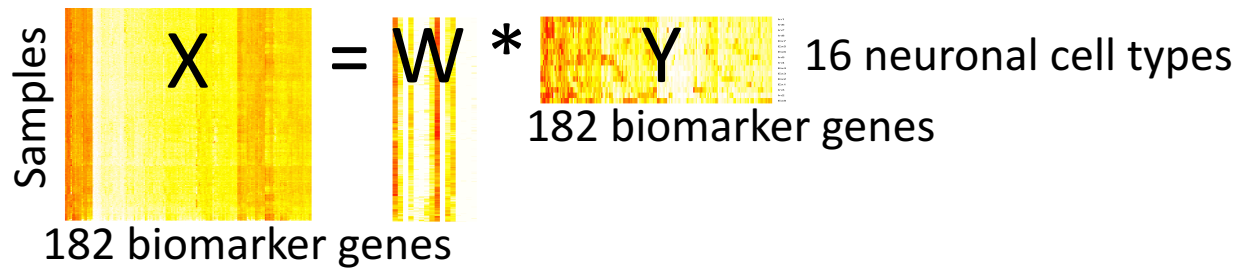


Brainspan

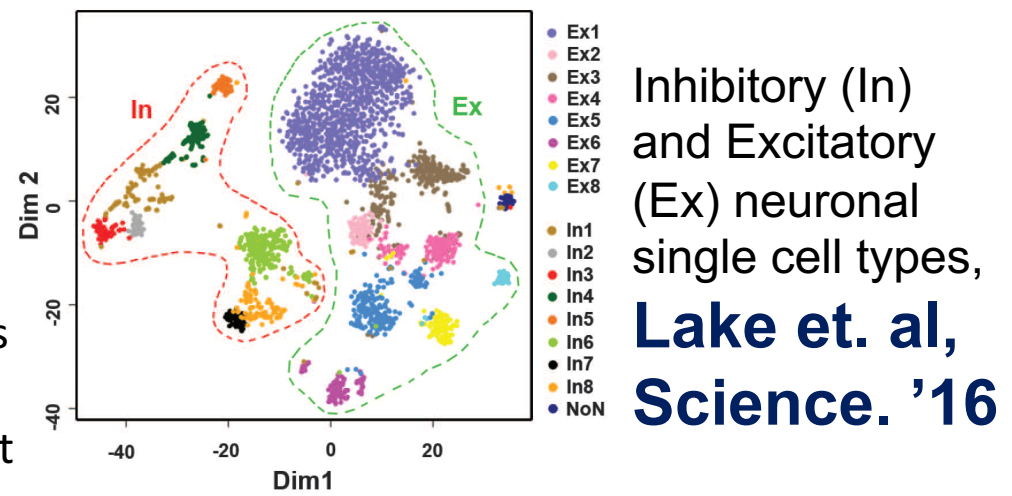
RCA



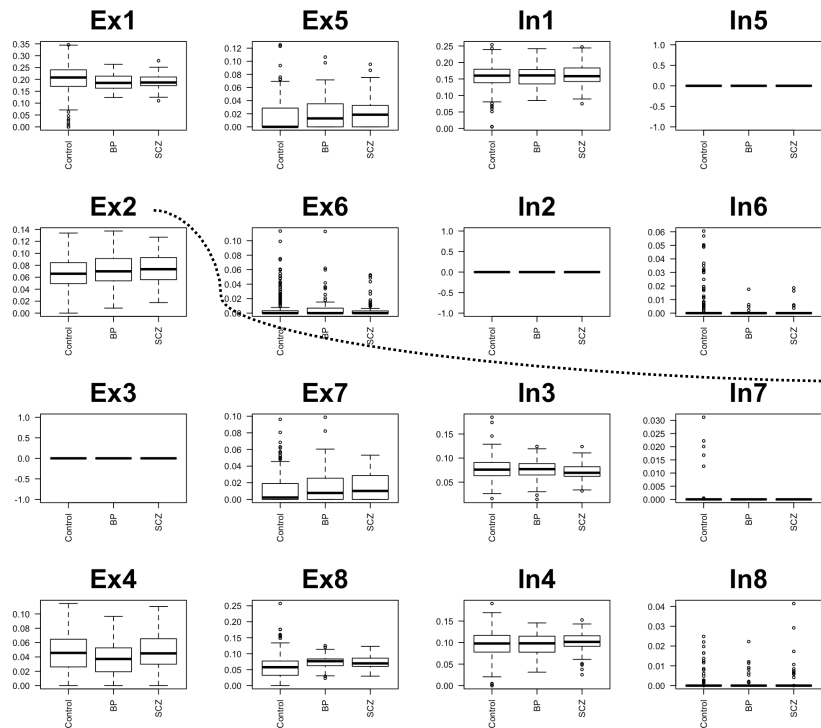
- Create a reference expression panel
 - Select ~1,400 genes with tissue specific expression
 - Calculating the median expression of all reference genes for all reference tissues
- For each sample, we project the gene expression values to the reference panel
- Finally we perform a PCA analysis using the projected values



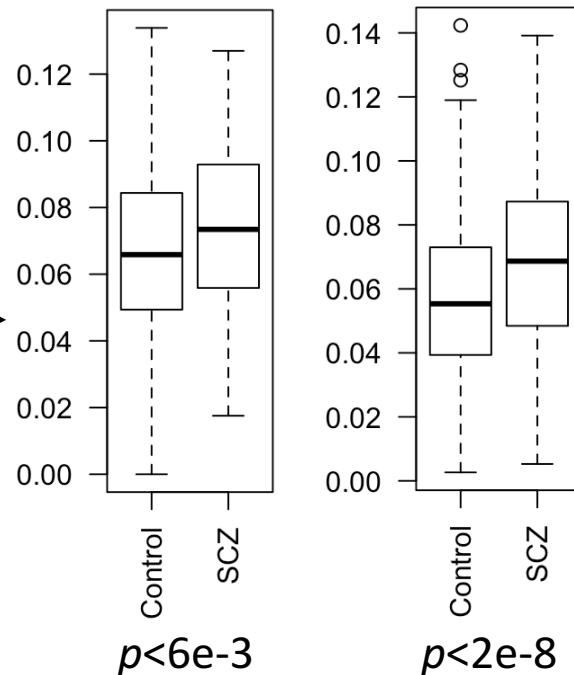
X – gene expression matrix of 182 neuronal cell type biomarker genes across brain cortex tissue; Y – avg. gene expression matrix of 182 biomarker genes across 16 neuronal cell types; W – neuronal cell type contribution coefficient matrix by deconvolution using non-negative least square



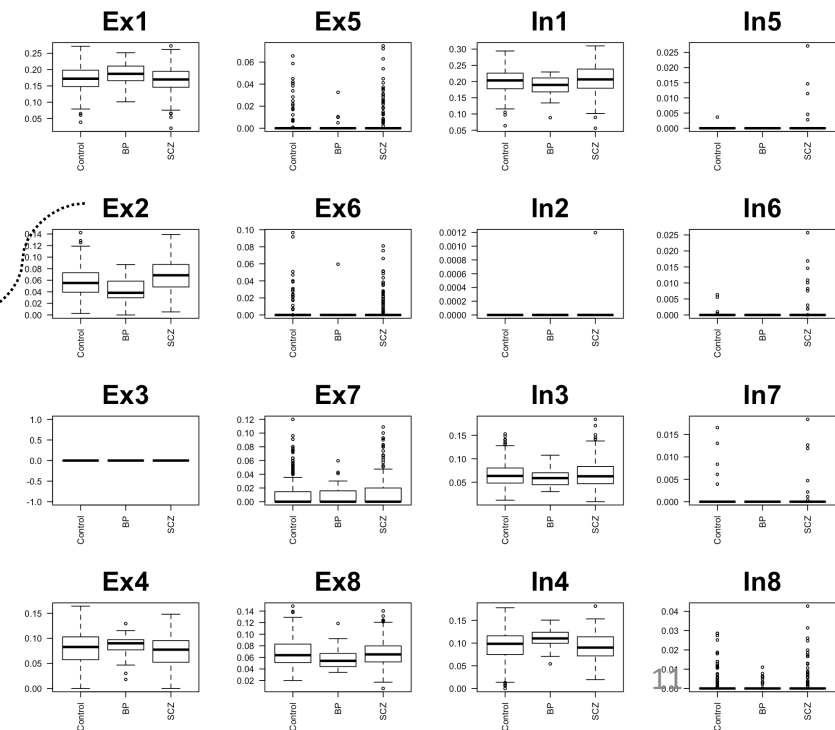
GVEX



Excitatory (Ex) neuronal type 2



CMC



$p < 2e-8$

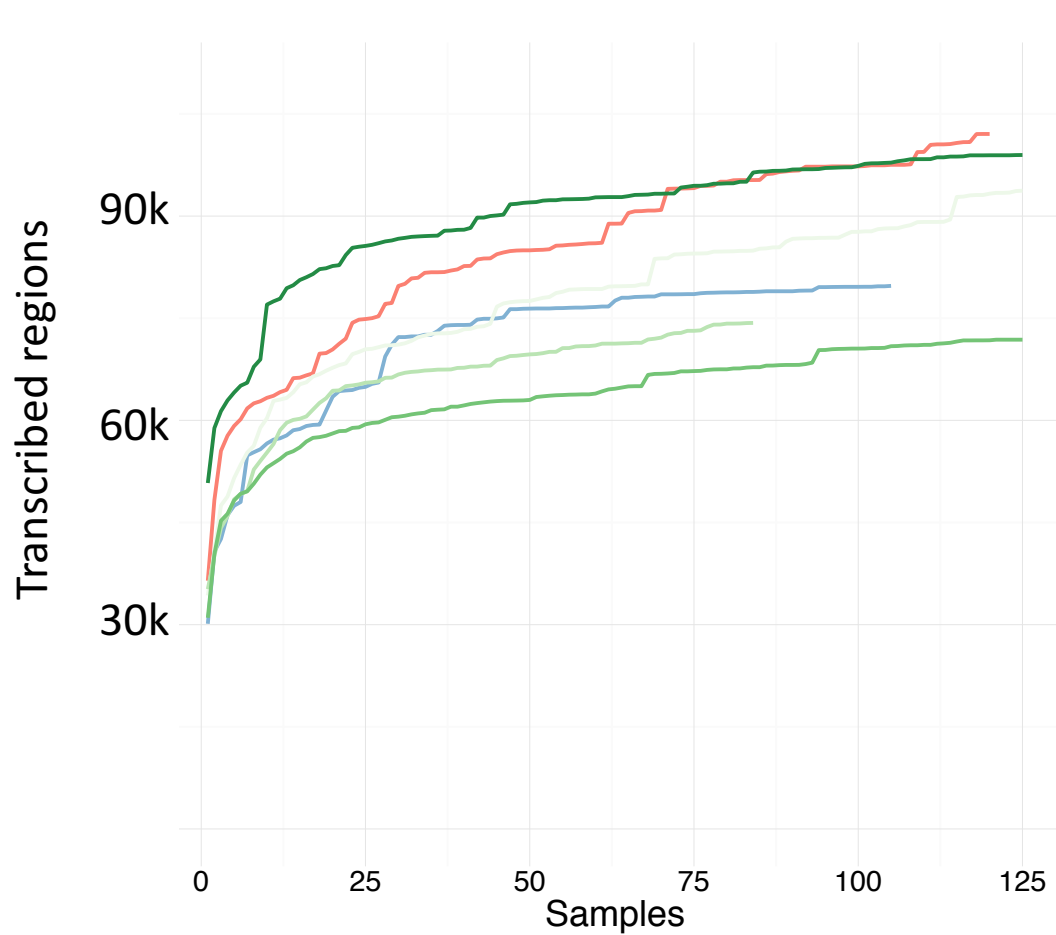
Single cell contribution coeffs for SCZ, BP, CTL brain cortex tissues

Capstone #4 – talk outline

- Resource
 - Brain enhancers, eQTLs, expression modules, networks &c
- Transcriptome analysis
 - 1) Spectral analysis of gene expression, unique brain expression patterns
 - 2) Non-coding transcription & retro-element activity, not atypical
- Integrative network models
 - Relating perturbations in gene expression to variants

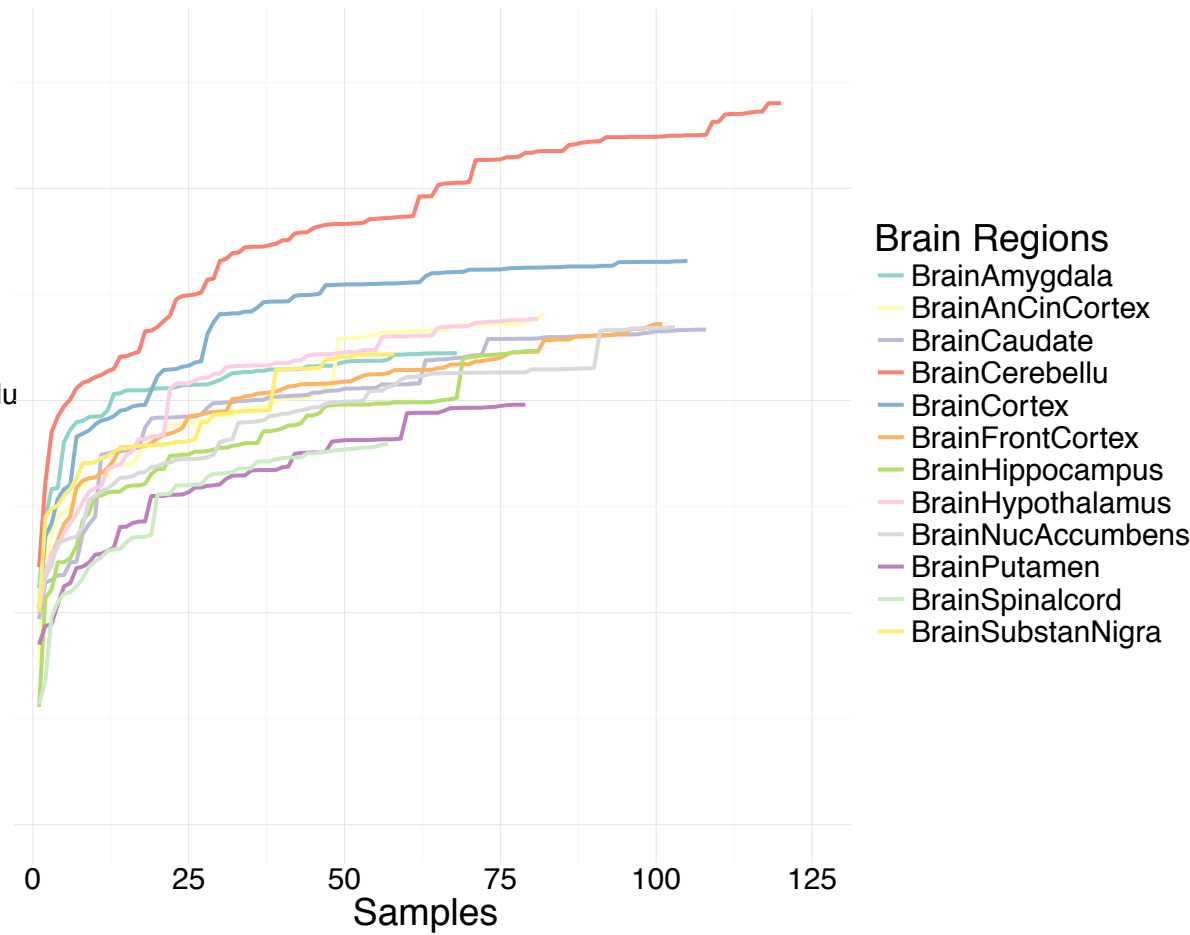
Transcriptome diversity, whole genome

90K regions = ~10% of genome



Tissues

- BrainCerebellu
- BrainCortex
- Lung
- Ovary
- Skin
- Testis



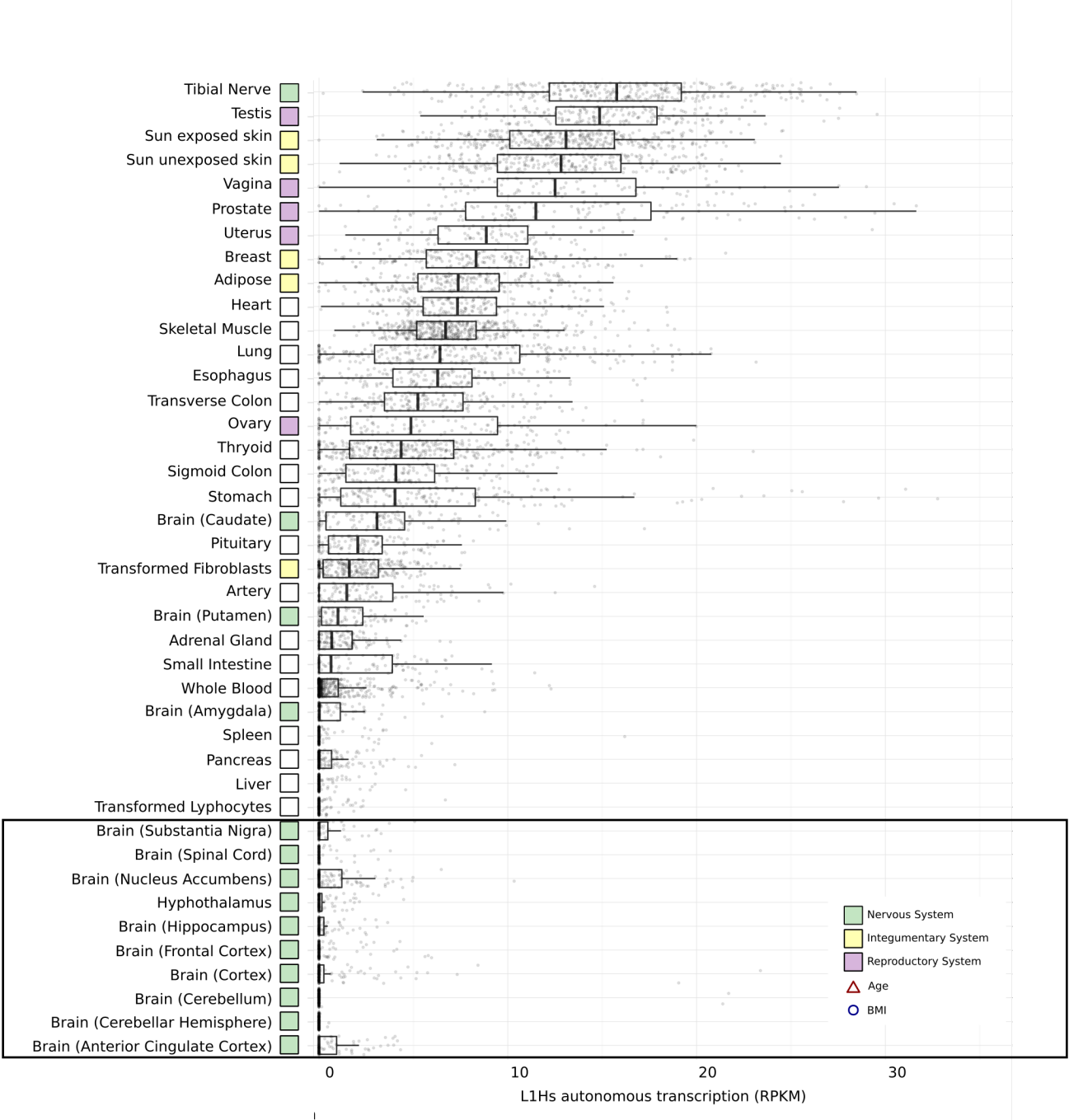
Brain Regions

- BrainAmygdala
- BrainAnCinCortex
- BrainCaudate
- BrainCerebellu
- BrainCortex
- BrainFrontCortex
- BrainHippocampus
- BrainHypothalamus
- BrainNucAccumbens
- BrainPutamen
- BrainSpinalcord
- BrainSubstanNigra

LINE-1 activity in the human brain

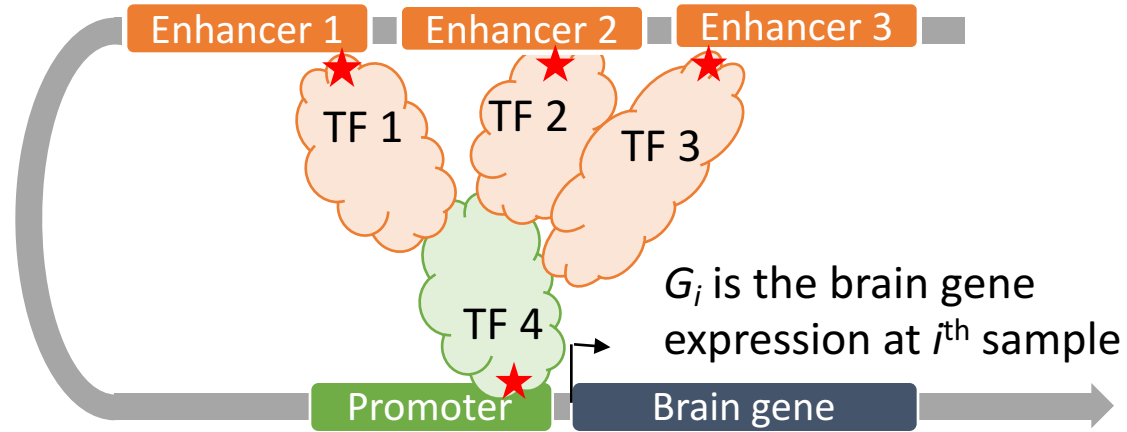
- LINE-1 has been shown to be active in the human brain
 - Creating somatic mosaicism across neurons (Muotri et. al.)
 - Suggesting that LINE-1 might contribute to neuronal plasticity
- We used **RNA-seq** data to uncover the activity of LINE-1 elements in **brain** regions
- However, ~30% of the human genome is derived from LINE-1 elements.
 - We developed Methods to **decouple pervasive** transcription from **autonomous** LINE-1 transcription.

Brain

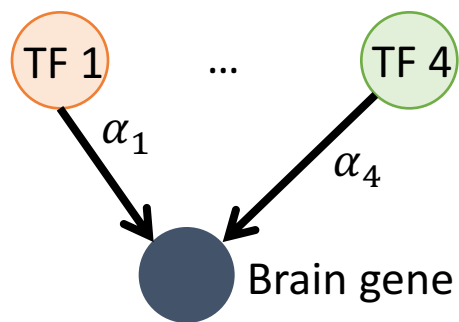
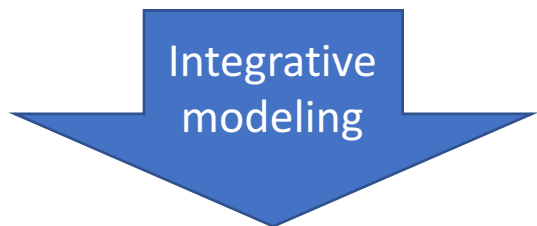


Capstone #4 – talk outline

- Resource
 - Brain enhancers, QTLs, expression modules, networks &c
- Transcriptome analysis
 - 1) Spectral analysis of gene expression, unique brain expression patterns
 - 2) Non-coding transcription & retro-element activity, not atypical
- Integrative network models
 - Relating perturbations in gene expression to variants

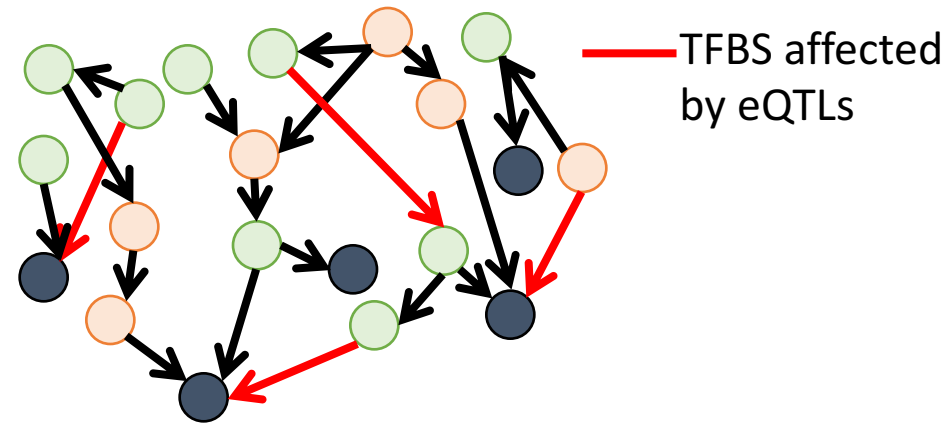
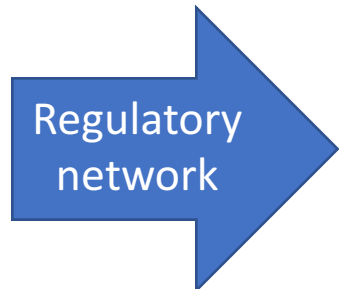


- ★ QTLs for brain gene ($E_{l,i}$ is l^{th} eQTL's genotype at i^{th} sample)
- Orange box Enhancers linking to brain gene
- Green box Brain gene promoter
- Cloud icon Transcription Factors (TFs) with TFBSs ($T_{k,i}$ is k^{th} TF expression at i^{th} sample)



$$G_i = \underbrace{\sum_{k=1}^n \alpha_k T_{k,i}}_{\text{Contributions of TFs}} + \underbrace{\sum_{l=1}^m \beta_l E_{l,i}}_{\text{eQTLs}} + \underbrace{\sum \gamma_{k,l} T_{k,i} \times E_{l,i}}_{\text{interactions of TFs and eQTLs}}$$

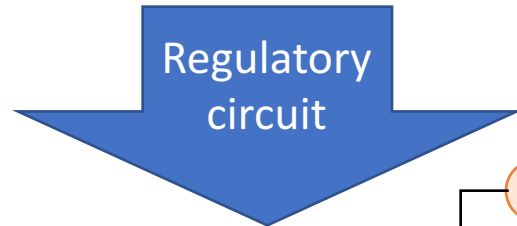
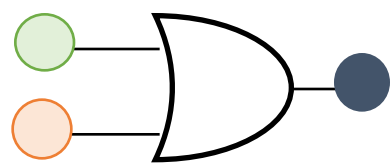
Contributions of TFs, eQTLs, interactions of TFs and eQTLs



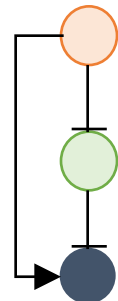
Network structure analysis



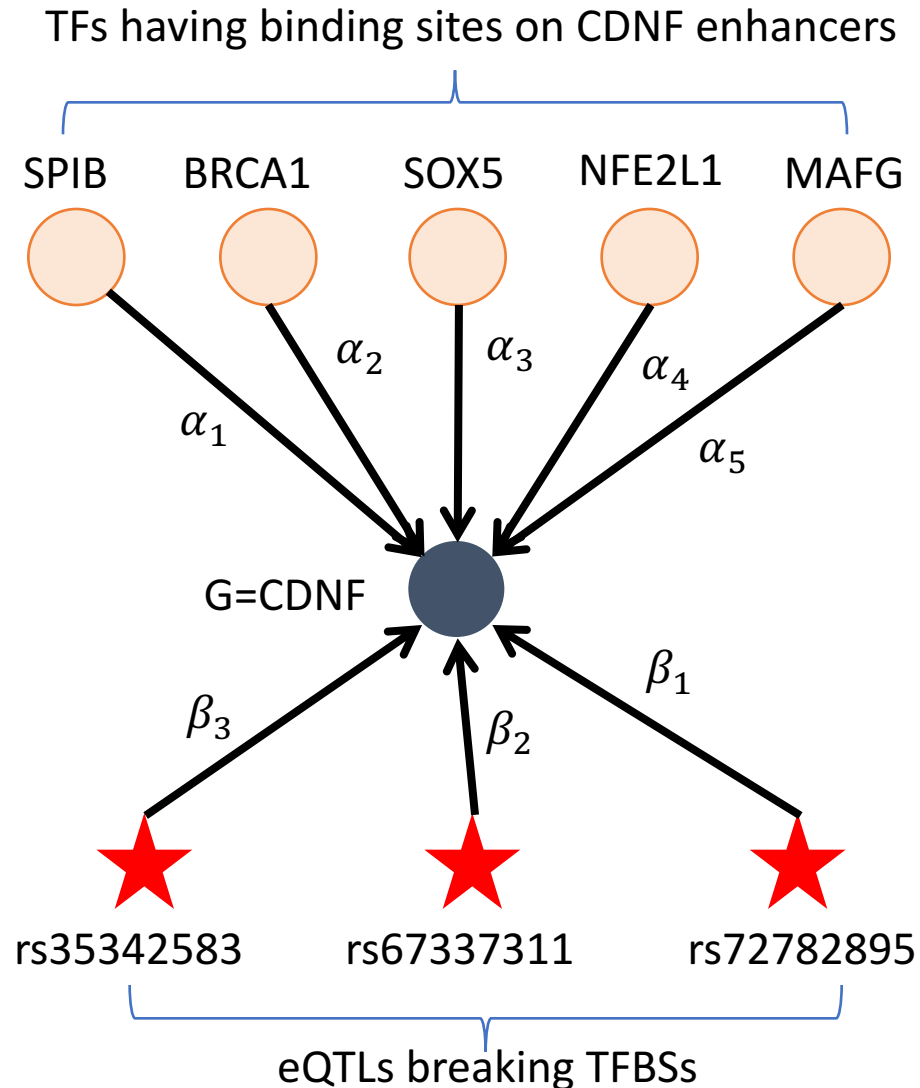
Input type	TF1	0	0	1	1
	TF4	0	1	0	1
Output	Brain gene	0	1	1	1



→ activation
 —| repression



Example: enhancer regulatory network for CDNF (Cerebral Dopamine Neurotrophic Factor)



$$G_i = \sum_{k=1}^n \alpha_k T_{k,i} + \sum_{l=1}^m \beta_l E_{l,i} + \sum \gamma_{k,l} T_{k,i} \times E_{l,i}$$

- SPIB $\alpha_1=0.18$
- BRCA1 $\alpha_2= 0.053$
- SOX5 $\alpha_3= -0.072$
- NFE2L1 $\alpha_4= 0.051$
- MAFG $\alpha_5= -0.389$
- rs72782895 $\beta_1= 0.066$
- rs67337311 $\beta_2= -0.025$
- rs35342583 $\beta_3= -0.032$

using CMC SCZ & CTRL gene expression and genotype data

Capstone #4 – talk outline

- Resource
 - Brain enhancers, QTLs, expression modules, networks &c
- Transcriptome analysis
 - 1) Spectral analysis of gene expression, unique brain expression patterns
 - 2) Non-coding transcription & retro-element activity, not atypical
- Integrative network models
 - Relating perturbations in gene expression to variants

PsychENCODE Capstone #4

Developing a Brain Genomics Resource via Integrating CommonMind, ENCODE, GTEx, and Roadmap

- **Shuang Liu**
- **Daifeng Wang**
- Fabio Navarro
- Declan Clarke
- Jonathan Warrell
- Prashant Emani
- Mengting Gu
- Timur Galeev
- Aparna Nathan
- Mark Gerstein
- **James Knowles**
- Pamela Sklar
- Mette Peters
- Gabriel Hoffman
- Laura Sloofman
- Andrew Jaffe
- Suhn Kyong Rhie
- Peggy Farnham