

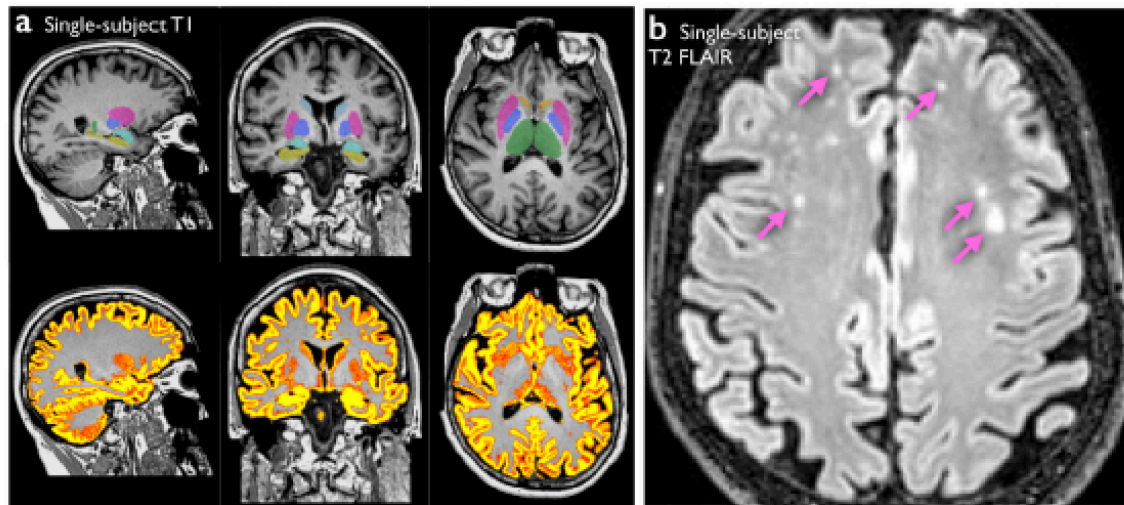
# **PHENOTYPES, GENOTYPES & VOXELS: A PLAYGROUND NEXT TO A NUCLEAR POWER PLANT**

Anqi Qiu

Department of Biomedical Engineering  
Clinical Imaging Research Centre  
National University of Singapore

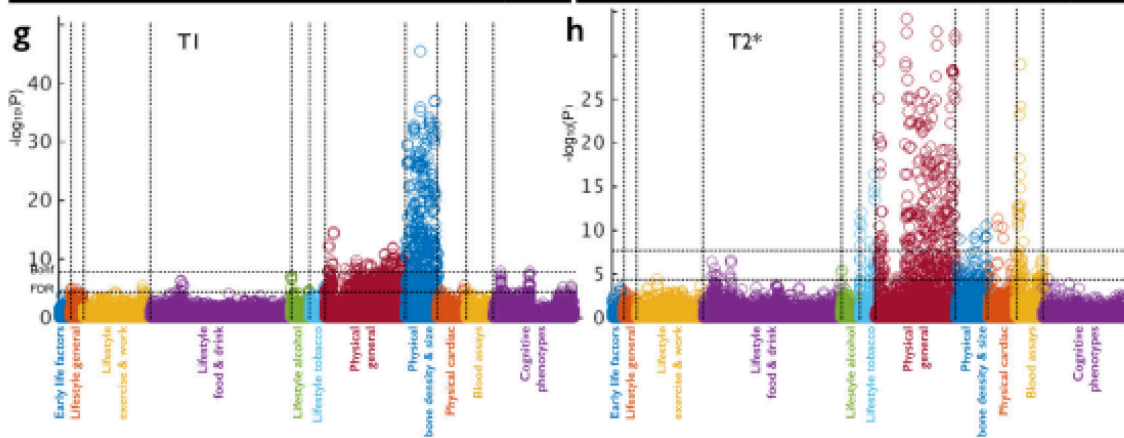


# ASSOCIATIONS BETWEEN VOXELS AND MULTIPLE PHENOTYPES



## UK BioBank

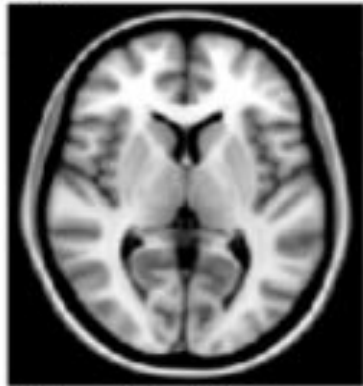
Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* 2016; 19(11): 1523-36.



# GENOME-WIDE ASSOCIATION STUDY (GWAS)

---

What are genetic markers related to brain imaging phenotypes?



Whole Brain

```
AGGAAGGTTGGAAACCCGGGGCCGGGGTTCC  
TTGGAAAACCAAGAAATCAATCCTAGGGGTC  
TGAGGTCTTTTAGTTAACCTCTGGCCCGG  
AATTTCCGGGGCCAGAAATCGAAAAATGGGAGT  
TAATCTTACCTCCTCTGTTAGCTTTCTCCA  
CCGGCTTTCCAAGATCTCTTTCCCGTGGTCG  
GCCTTGGAGGACCTTAGAGGGGACTGGCCA  
AAAAAATTTCTGAATCCCTTCTGGCTAGTTG  
GAGAAAGGCCAGTCCGTTGATCGGAGGGGAA  
CAGCTTTTTCCTTTTTCCTTTCCTTTC
```

Individual SNPs



# OUTLINE

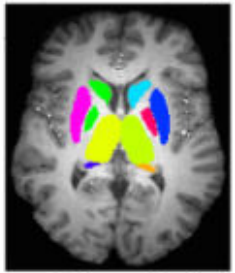
---

- BIG DATA
  - 1) BIG CONFOUND TROUBLE IN GENOME-WIDE ASSOCIATION STUDIES
  - 2) POSSIBLE SOLUTION TO ESTIMATE CONFOUNDING EFFECTS
- SMALL DATA
  - 1) CHALLENGES IN GENOME-WISE ASSOCIATION STUDIES WHEN THE SAMPLE SIZE IS SMALL;
  - 2) POSSIBLE SOLUTIONS: HAPLOTYPE ANALYSIS; POLYGENIC RISK SCORE; GENE-SET ANALYSIS.
- GENE EXPRESSION FROM POST-MORTEM BRAIN





# GENOME-WIDE ASSOCIATION STUDY (GWAS)



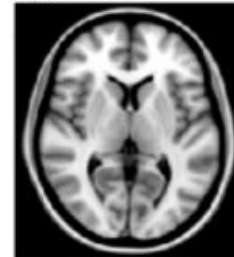
Subcortical Volumes

```
AGGAAGGTTGGAAACCCCGGGGCCGGGGTTCC
TTGGAAAACCAAGAAATCAATCCTAGGGGTC
TGAGGTCTTTTAGTTAACCTCTGGCCCCGG
AATTTCCGGGGCCAGAATCGAAAAATGGGAGT
TAATCTTACCTCCTCTGTTAGCTTTCTCCAA
CCGGCTTTCCAGATCTCTTTCCCGTGGTCCG
GCCTTGGAGGACCTTAGAGGGGACTGGCCAA
AAAAAATTTCTGAATCCCTTCTGGCTAGTTG
GAGAAGGCCAGTCGGTTGATCGGAGGGGAAA
CAGCTTTTTTTCCTTTTTCCTTTCCTTTC
```

Individual SNPs

Associations of Genetic Variants and Structural Volumes

Hibar DP, *et al.*, "Common genetic variants influence human subcortical brain structures." *Nature* 2015.



Whole Brain

```
AGGAAGGTTGGAAACCCCGGGGCCGGGGTTCC
TTGGAAAACCAAGAAATCAATCCTAGGGGTC
TGAGGTCTTTTAGTTAACCTCTGGCCCCGG
AATTTCCGGGGCCAGAATCGAAAAATGGGAGT
TAATCTTACCTCCTCTGTTAGCTTTCTCCAA
CCGGCTTTCCAGATCTCTTTCCCGTGGTCCG
GCCTTGGAGGACCTTAGAGGGGACTGGCCAA
AAAAAATTTCTGAATCCCTTCTGGCTAGTTG
GAGAAGGCCAGTCGGTTGATCGGAGGGGAAA
CAGCTTTTTTTCCTTTTTCCTTTCCTTTC
```

Individual SNPs

Associations of Genetic Variants and Voxel-Wise Imaging Measures

Huang M, *et al.*, "Fast voxelwise genome wide association analysis of large-scale imaging genetic data", *Neuroimage*, 2015.

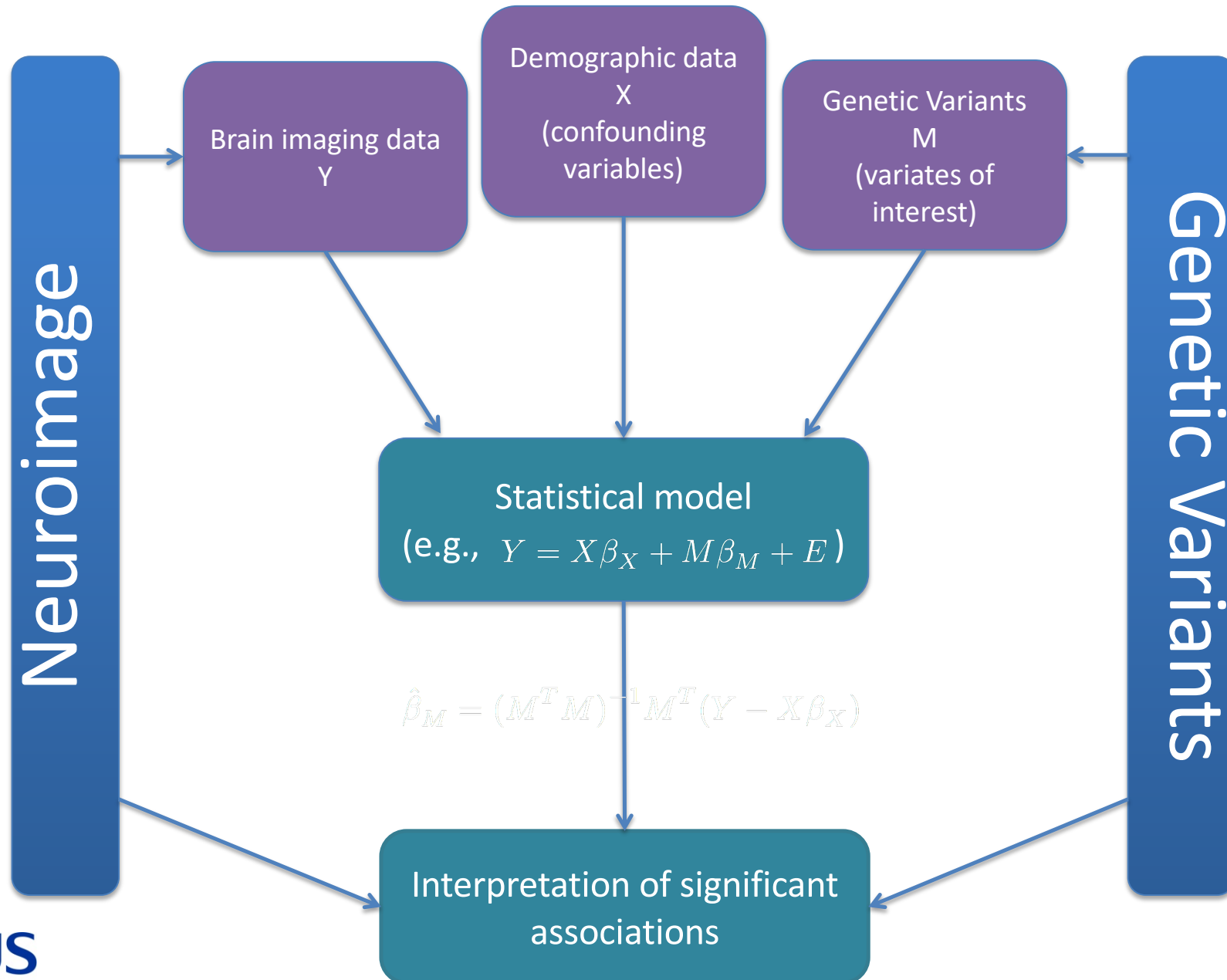


“CURRENTLY, THE WORKHORSE TOOLS OF GWAS ON NEUROIMAGING ARE “MASS UNIVARIATE”, WHERE A REGRESSION MODEL IS FIT SEPARATELY AT EACH IMAGING MEASURE.”

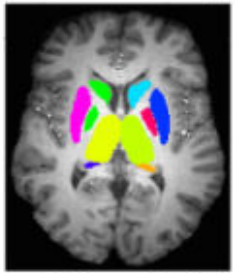
Smith SM, Nichols TE. (2018): Statistical Challenges in "Big Data" Human Neuroimaging. Neuron 97(2):263-268.



# MASS-UNIVARIATE ANALYSIS VIA REGRESSION



# GENOME-WIDE ASSOCIATION STUDY (GWAS)

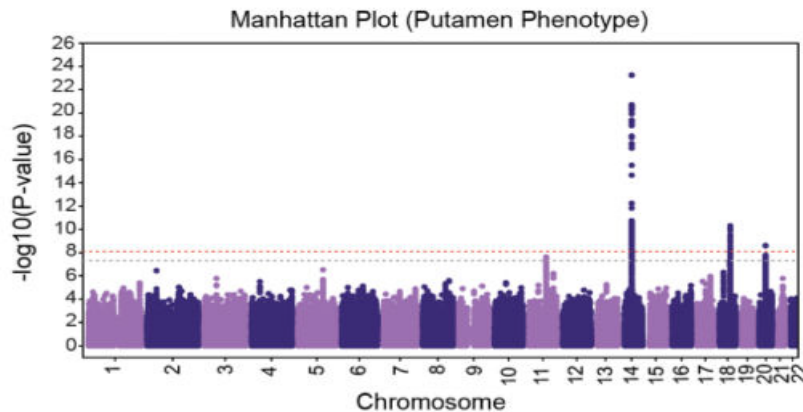


Subcortical Volumes

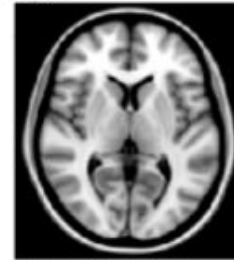
```
AGGAAGGTTTGGAAACCCCGGGGCCGGGGTTCC
TTGGAAAACCAAGAAATCAATCCTAGGGGTC
TGAGGTCTTTTAGTTAACCCCTCTGGCCCCGG
AATTTCCGGGCCAGAATCGAAAATGGGAGT
TAATCTTACCTCCTCTGTAGCTTTCTCCAA
CCGGCTTTCCAAAGATCTCTTTCCCGTGGTCCG
GCCTTGGAGGACCTTAGAGGGGACTGGCCAA
AAAAAATTTCTGAATCCCTTCTGGCTAGTTG
GAGAAGGCCAGTCGGTTGATCGGAGGGGAAA
CAGCTTTTTTTCCTTTTTCCTTTCCTTTCCTTTC
```

Individual SNPs

Associations of Genetic Variants and Structural Volumes



Hibar DP, *et al.*, "Common genetic variants influence human subcortical brain structures." *Nature* 2015.



Whole Brain

```
AGGAAGGTTTGGAAACCCCGGGGCCGGGGTTCC
TTGGAAAACCAAGAAATCAATCCTAGGGGTC
TGAGGTCTTTTAGTTAACCCCTCTGGCCCCGG
AATTTCCGGGCCAGAATCGAAAATGGGAGT
TAATCTTACCTCCTCTGTAGCTTTCTCCAA
CCGGCTTTCCAAAGATCTCTTTCCCGTGGTCCG
GCCTTGGAGGACCTTAGAGGGGACTGGCCAA
AAAAAATTTCTGAATCCCTTCTGGCTAGTTG
GAGAAGGCCAGTCGGTTGATCGGAGGGGAAA
CAGCTTTTTTTCCTTTTTCCTTTCCTTTCCTTTC
```

Individual SNPs

Associations of Genetic Variants and Voxel-Wise Imaging Measures



Statistical approaches to deal with correction for a large number of statistical tests.

Huang M, *et al.*, "Fast voxelwise genome wide association analysis of large-scale imaging genetic data", *Neuroimage*, 2015.



# CHALLENGE IN GWAS

---

“BIG DATA: BIG CONFOUND TROUBLE.”

Smith SM, Nichols TE. (2018): Statistical Challenges in "Big Data" Human Neuroimaging. *Neuron* 97(2):263-268.



# BIG CONFOUND TROUBLE

$$Y = X\beta_X + M\beta_M + E$$

Assumption:  $X$  is a complete set of confounding variables

$$\hat{\beta}_M = (M^T M)^{-1} M^T (Y - X\beta_X)$$

$$\text{Var}(\hat{\beta}_M) = \sigma^2 (M^T M)^{-1}$$

- Inflate residual
- Inflate variance estimates
- Deflate statistics
- Make hypothesis tests conservative

$Cor(M, Z)$   
 $Z$  are missing covariates

Positive or negative bias for the association between  $M$  and  $Y$ .



# POSSIBLE SOLUTION

---

$$Y = X\beta_X + M\beta_M + Z\beta_Z + E$$

prior knowledge

unknown knowledge

Software is available at <https://bieqa.github.io/imaginggenetics.html>

Bryan Guillaume, Changqing Wang, Joann Poh, Mo Jun Shen, Mei Lyn Ong, Pei Fang, Neerja Karnani, Michael Meaney, Anqi Qiu, "Improving mass-univariate analysis of neuroimaging data by modelling important unknown covariates: application to Epigenome-Wide Association Studies", *Neuroimage*, 173:57-71, 2018.





# EXAMPLE: APPLICATION TO EPIGENOME-WIDE ASSOCIATION STUDY

---

How does our body remember what happened to us **previously**, and how does that affect **brain development**?

Imaging and Epigenetic Datasets:

- 114 neonatal brain DTI datasets
- 42372 methylation loci

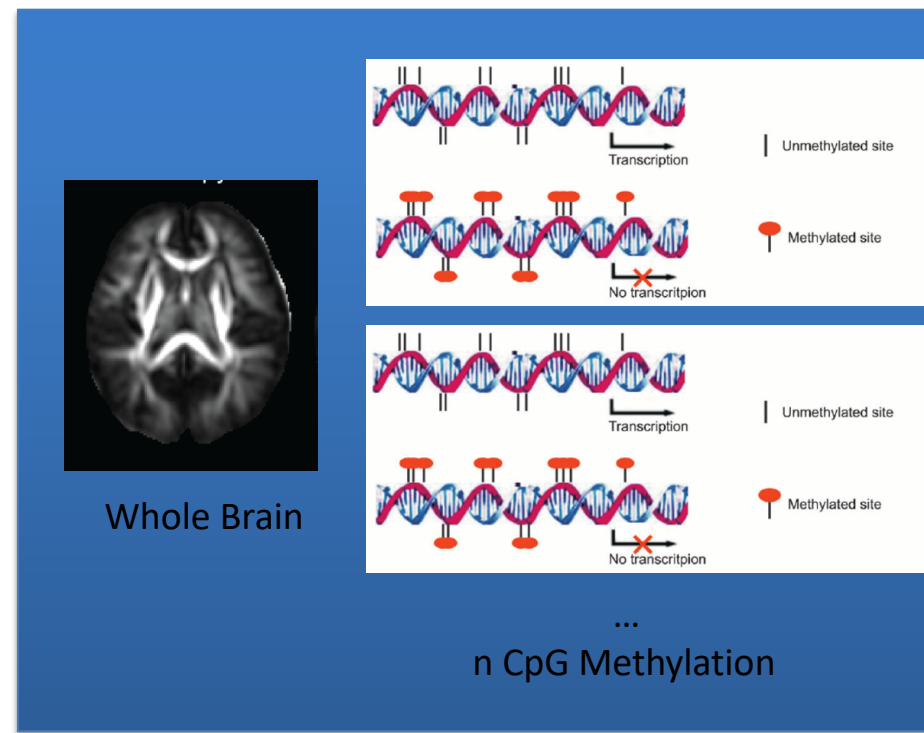


# EPIGENOME-WIDE ASSOCIATION STUDY ON NEONATAL BRAIN

How does our body remember what happened to us **previously**, and how does that affect **brain development**?

Imaging and Epigenetic Datasets:

- 114 neonatal brain DTI datasets
- 42372 epigenetic loci
- Covariates: gestational age at MRI, gender, birth weight.



$$Y = X\beta_X + M\beta_M + Z\beta_Z + E$$

multiple comparisons  
across the whole brain and  
across methylation markers



# SIGNIFICANT ASSOCIATIONS BETWEEN CG19641625 AND VOXELS IN THE LEFT VENTROLATERAL THALAMUS

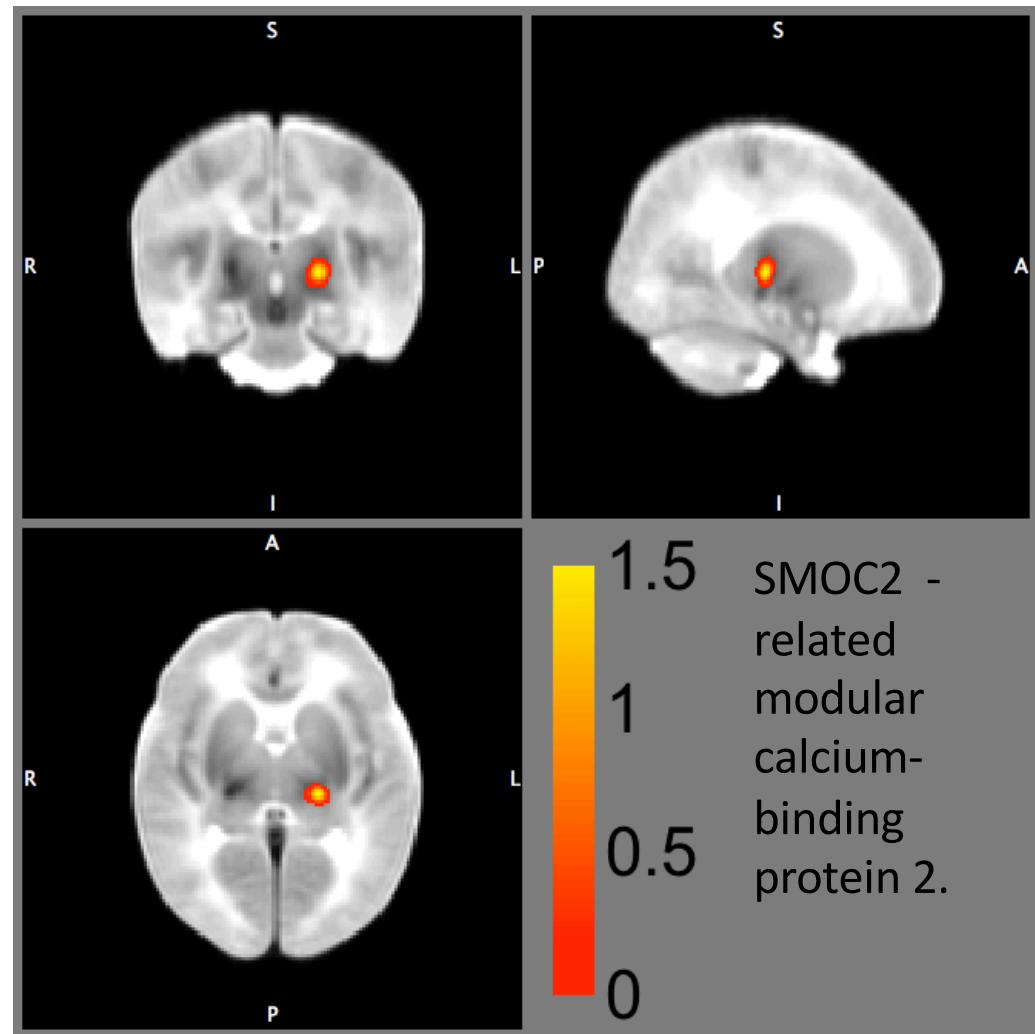
How does our body remember what happened to us **previously**, and how does that affect **brain development**?

## Imaging and Epigenetic Datasets:

- 114 neonatal brain DTI datasets
- 42372 epigenetic loci
- Covariates: gestational age at MRI, gender, birth weight.

## Findings:

- Significant association between cg19641625 and FA in the left ventrolateral thalamus



$-\log_{10}(\text{FWER-corrected p-value})$  image obtained for the methylation locus cg19641625



# SIGNIFICANT ASSOCIATIONS BETWEEN CG19641625 AND VOXELS IN THE LEFT VENTROLATERAL THALAMUS

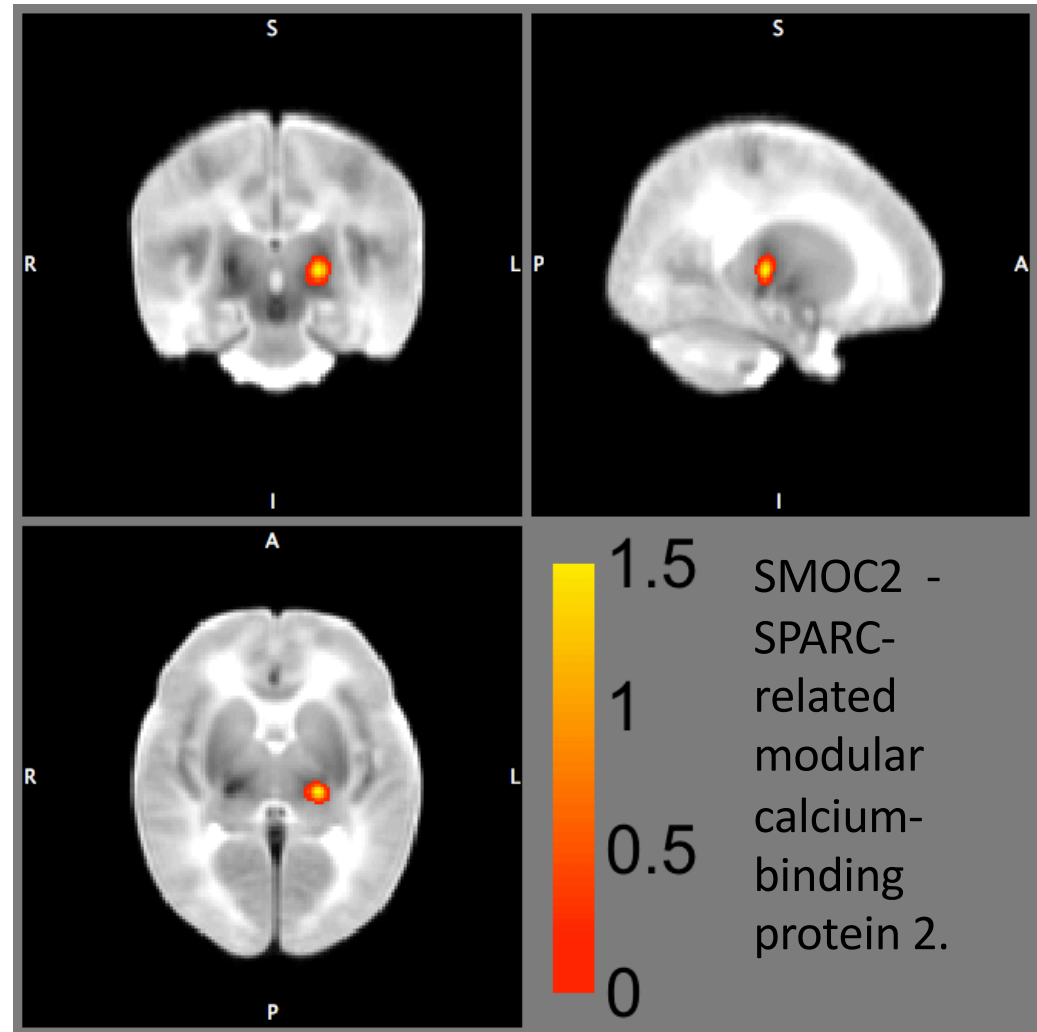
How does our body remember what happened to us **previously**, and how does that affect **brain development**?

## Imaging and Epigenetic Datasets:

- 114 neonatal brain DTI datasets
- 42372 methylation loci
- Covariates: gestational age at MRI, gender, birth weight.

## Findings:

- Significant association between cg19641625 and FA in the left ventrolateral thalamus
- **2 estimated missing confounding variables – highly correlated with birth length and total brain volume**



$-\log_{10}(\text{FWER-corrected p-value})$  image obtained for the methylation locus cg19641625



# OTHER CHALLENGES IN GWAS

---

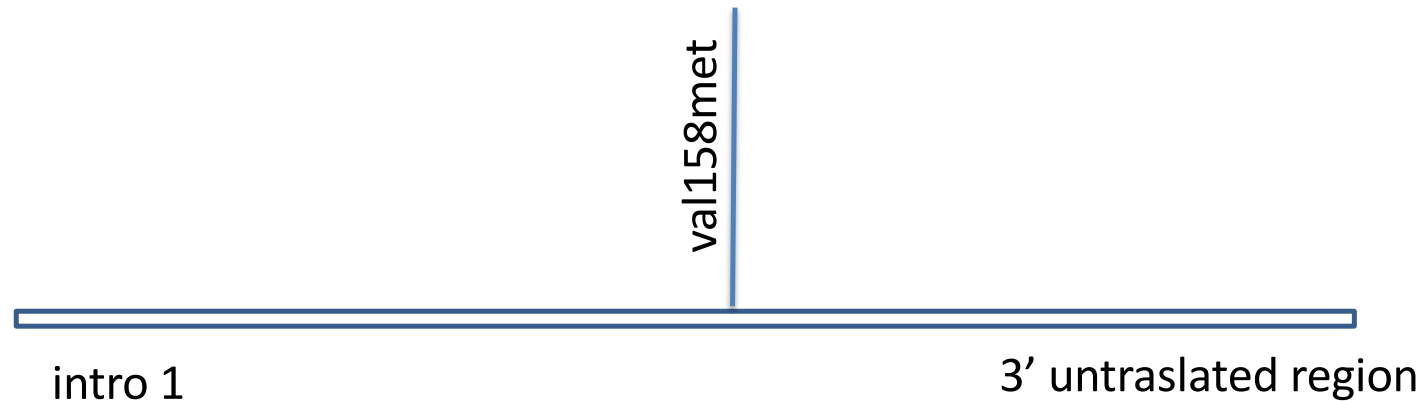
- COMMON GENETIC VARIANTS ONLY ACCOUNT FOR A SMALL FRACTION OF THE HERITABILITY OF BRAIN PHENOTYPES.



# COMMON GENETIC VARIANTS ACCOUNT FOR SMALL HERITABILITY

---

## COMT VAL158MET



COMT regulates dopamine signaling in the prefrontal cortex. A functional SNP of the COMT gene (*val158met*) leads to an amino acid change from valine (*val*) to methionine (*met*) at position 158. The *val* allele is suggested as a predominant factor that determines higher COMT activity in the prefrontal cortex, which presumably leads to lower synaptic dopamine levels through enhanced degradation.

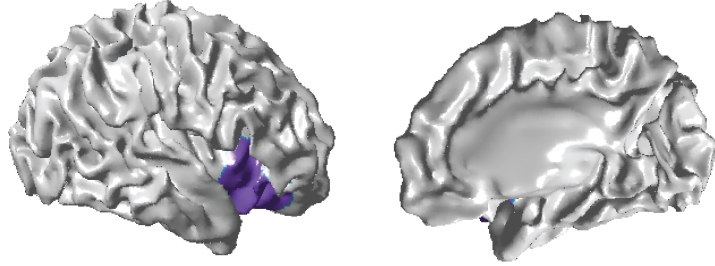


# COMT VAL158MET MODULATES ASSOCIATIONS OF ANTENATAL MATERNAL ANXIETY AND NEONATAL CORTICAL MORPHOLOGY

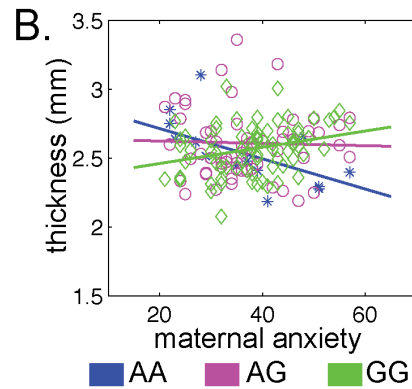
A. rs4680 (val158met)

Right Lateral View

Right Medial View



corrected p-value 0 0.01 0.02 0.03 0.04 0.05



Anqi Qiu et al., "COMT Haplotypes Modulate Associations of Antenatal Maternal Anxiety and Neonatal Cortical Morphology", *American Journal of Psychiatry*, 2014.

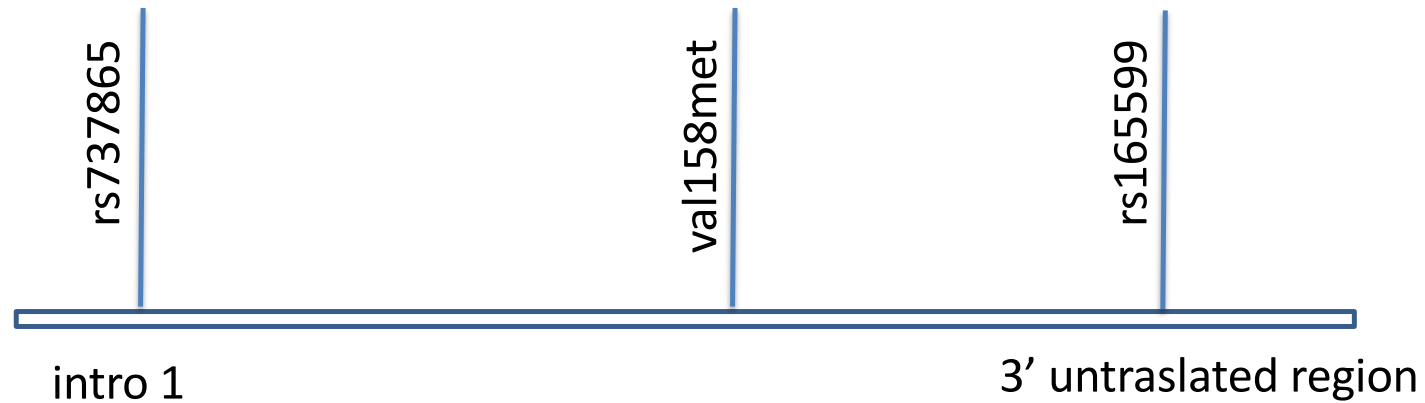




# HAPLOTYPE ANALYSIS

---

## COMT VARIANTS

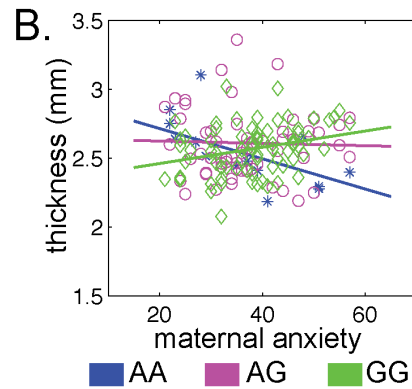
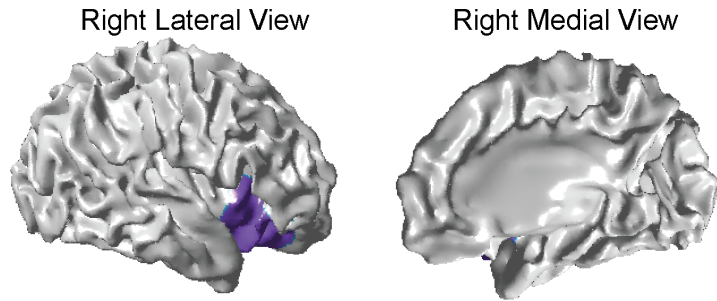


Sequence variants lying in other loci within the *COMT* gene affect gene function, ultimately resulting in altered enzyme activity and adding complexity to the functional and clinical implications of *COMT* variations.

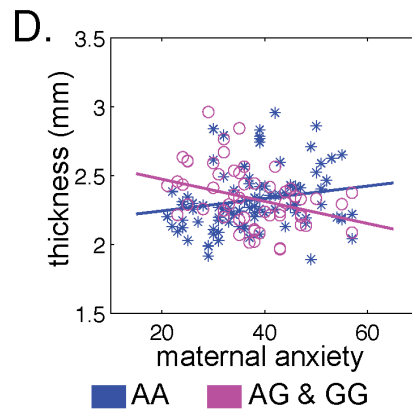
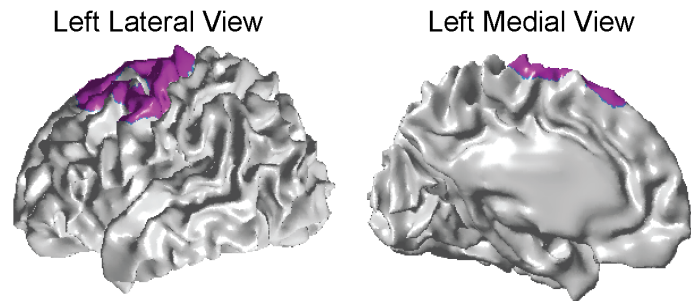


# COMT SNPs MODULATE ASSOCIATIONS OF ANTENATAL MATERNAL ANXIETY AND NEONATAL CORTICAL MORPHOLOGY

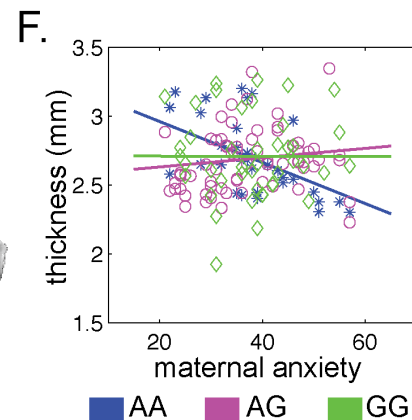
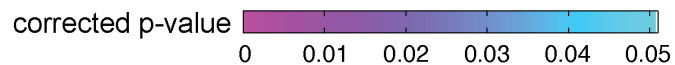
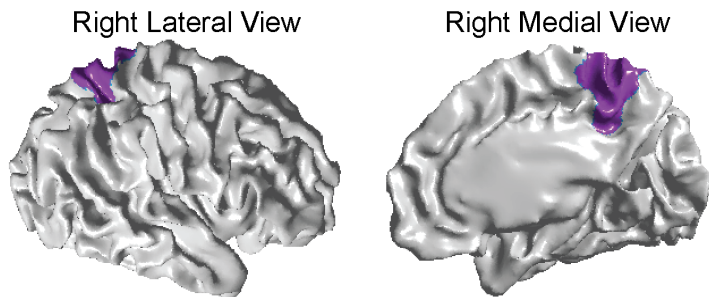
A. rs4680 (val158met)



C. rs737865



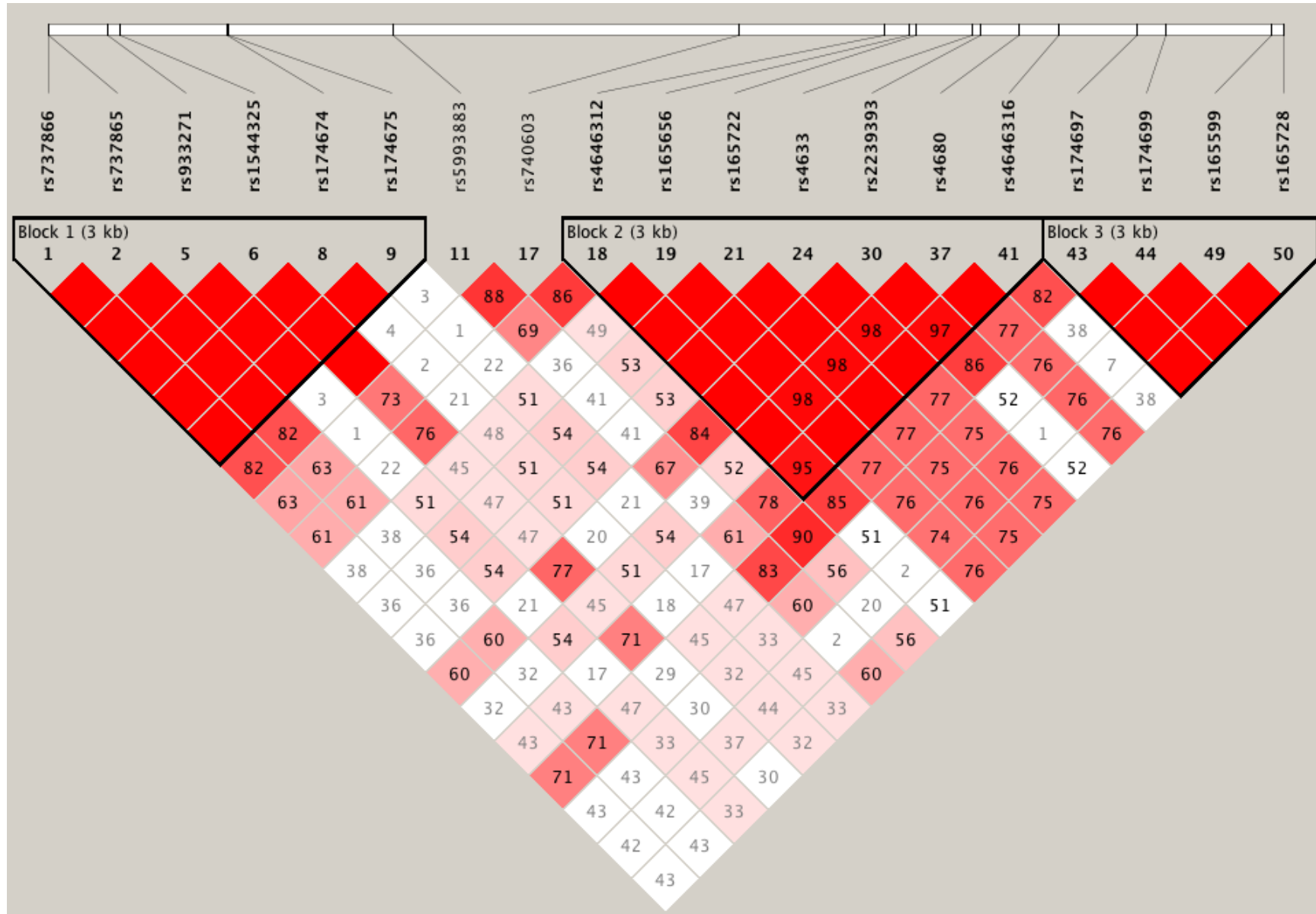
E. rs165599



Anqi Qiu et al., "COMT Haplotypes Modulate Associations of Antenatal Maternal Anxiety and Neonatal Cortical Morphology", *American Journal of Psychiatry*, 2014.

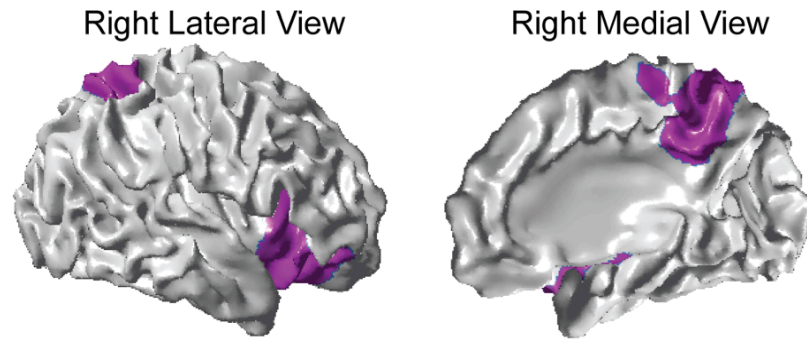


# COMT HAPLOTYPE



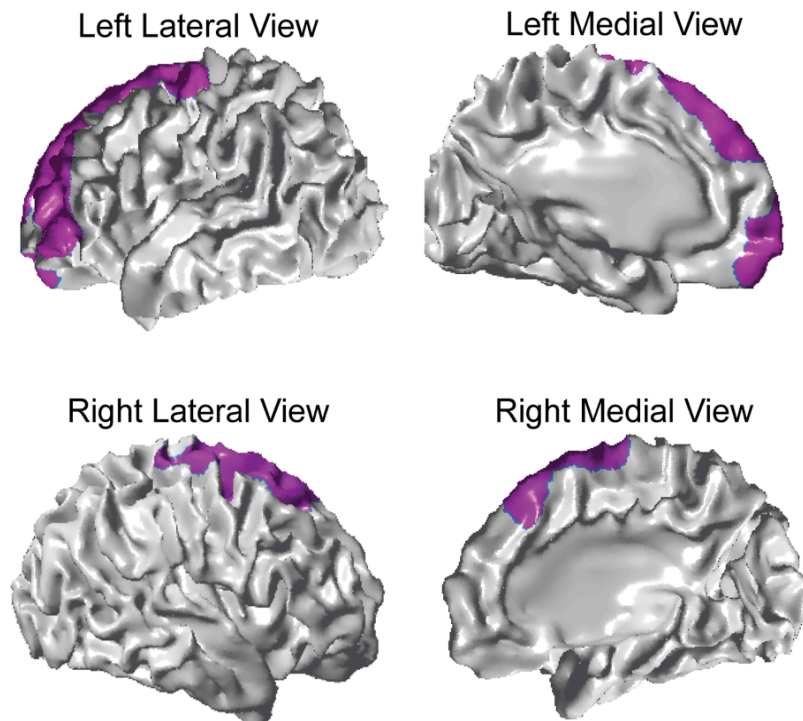
# COMT HAPLOTYPES MODULATE ASSOCIATIONS OF ANTENATAL MATERNAL ANXIETY AND NEONATAL CORTICAL MORPHOLOGY

## A. AGG Haplotype



Rs737865-val158met-rs165599

## B. GAA Haplotype



Anqi Qiu et al., "COMT Haplotypes Modulate Associations of Antenatal Maternal Anxiety and Neonatal Cortical Morphology", *American Journal of Psychiatry*, 2014.



# HYPLOTYPE ANALYSIS

---

- COMMON GENETIC VARIANTS ONLY ACCOUNT FOR A SMALL FRACTION OF THE HERITABILITY OF BRAIN PHENOTYPES.



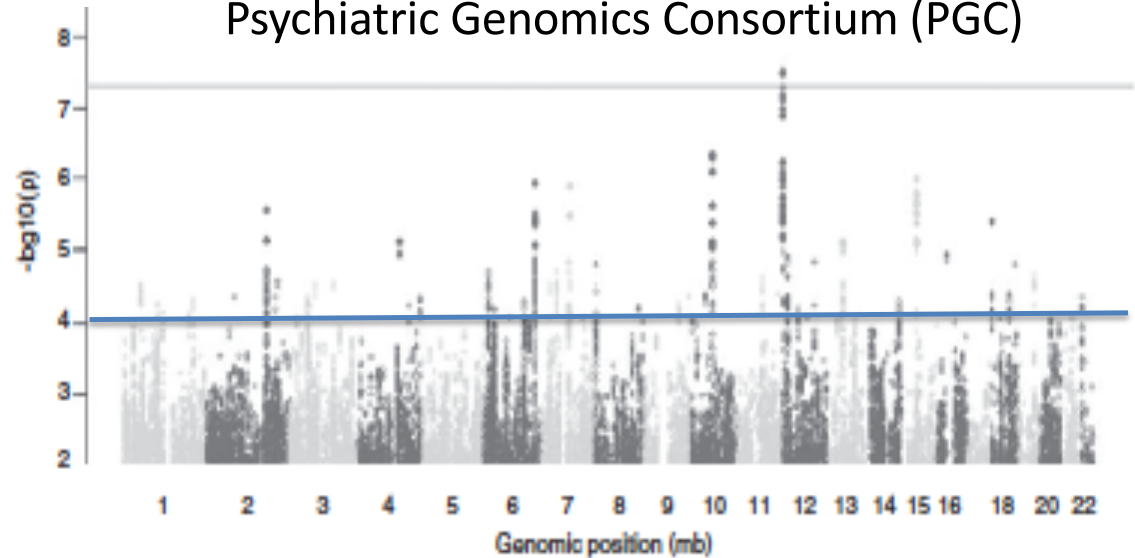
# POLYGENIC RISK SCORE

How does gene and antenatal maternal depressive symptoms affect **fetal brain development?**

Imaging and Genetic Datasets:

- 168 neonatal brain images
- Genotype data of 168 neonates

the discovery depression sample from Psychiatric Genomics Consortium (PGC)

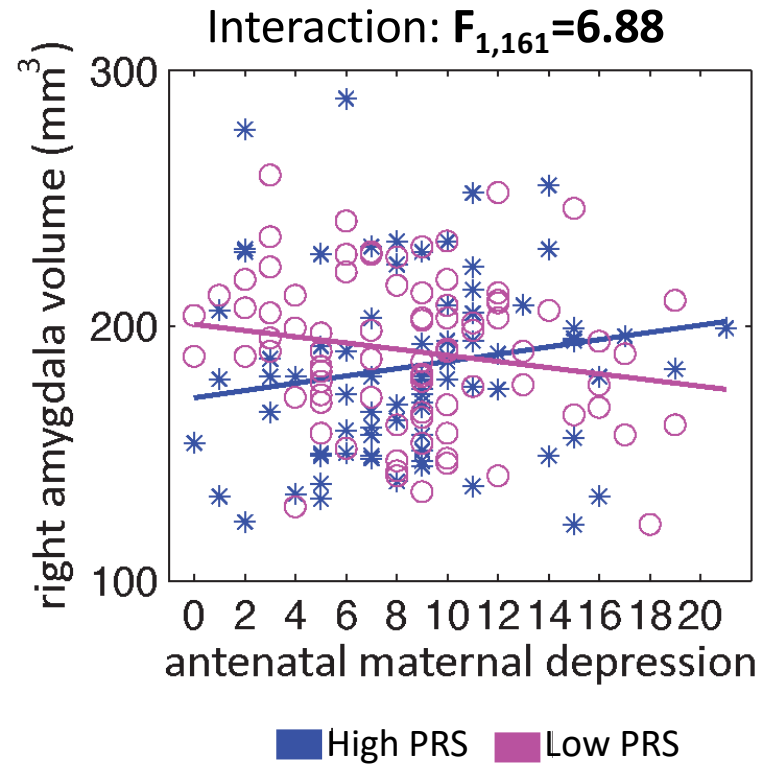


Individual SNPs

```
A G G A A G G T T G G A C C C C G G G C C G G G G T T C C  
T T G G A A A C C A A G A A A T C A A T C C T A G G G G T C  
T G A G G T C T T T T A G T T A C C C T C T G G C C C G G  
A A T T T C G G G G C C A G A A T C G A A A A T G G G A G T  
T A A T C T T A C C T C C T C T G T T A G C T T T C T G C C A  
C C G G C T T T C C A A G A T C T C T T T C C C G T G G T C G  
G C C T T G G A G G A C C T T A G A G G G G A C T G G C C A A  
A A A A A A T T T C T G A A T C C C T T C T G G C T A G T T G  
G A G A A G G C C A G T C G G T T G A T C G G A G G G G A A A  
C A G C T T T T T C C C T T T C C C A A A A C C T T C
```



# POLYGENIC RISK MODULATES ASSOCIATIONS OF PRENATAL MATERNAL DEPRESSION AND RIGHT AMYGDALA OF NEONATES



Anqi Qiu, Mojun Shen, Claudia Buss, Yap-Seng Chong, et al., "Effects of Antenatal Maternal Depressive Symptoms and Socio-economic Status on Neonatal Brain Development are Modulated by Genetic Risk", *Cerebral Cortex*, 27(5):3080-3092, 2017.







# OTHER CHALLENGES IN GWAS

---

- COMMON GENETIC VARIANTS ONLY ACCOUNT FOR A SMALL FRACTION OF THE HERITABILITY OF BRAIN PHENOTYPES.
- ``MISSING'', OR PERHAPS ``HIDDEN'', HERITABILITY MAY BE ACCOUNTED FOR BY RARE VARIANTS, WHICH ARE NOT EVEN CONSIDERED IN GWAS DUE TO THE LACK OF STATISTICAL POWER.
- RARE VARIANTS CAN ONLY BE IDENTIFIED USING GWAS WHEN A SAMPLE SIZE IS EXTREMELY LARGE, WHICH IS DIFFICULT TO ACHIEVE IN IMAGING GENETICS.



# A MIXED EFFECT MODEL FOR GENE SET AND ENVIRONMENT INTERACTION (MIXGE)

$$g\{E(y_i)\} = X_i^T \beta_x + e_i \beta_e + G_i^T W \pi^{(1)} + e_i G_i^T W \pi^{(2)} + e_i G_i^T \delta$$

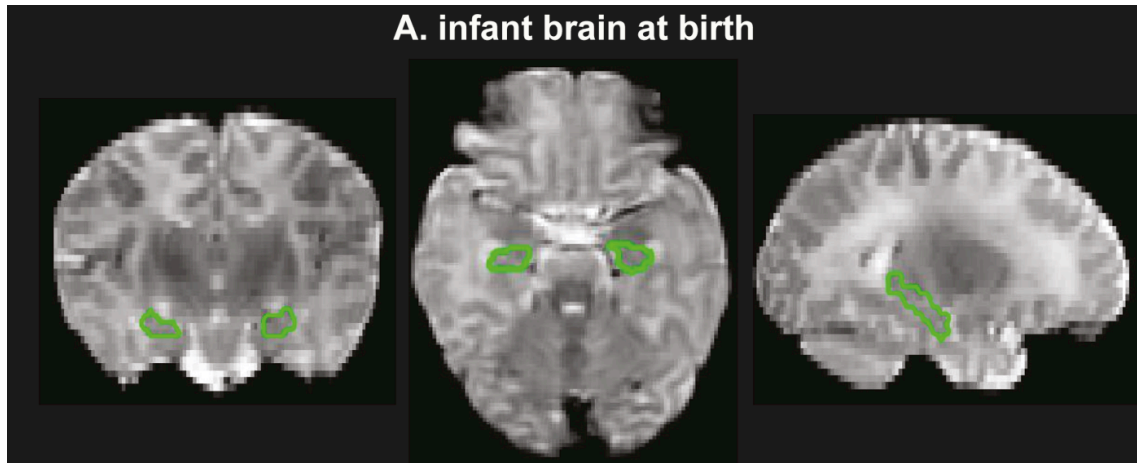
Diagram illustrating the components of the Mixed Effect Model for Gene Set and Environment Interaction (MIXGE):

- $X_i^T \beta_x$ : covariates
- $e_i \beta_e$ : environmental factor
- $G_i^T W \pi^{(1)}$ : gene set
- $e_i G_i^T W \pi^{(2)}$ : genetic score
- $e_i G_i^T \delta$ : interaction term



# GENE SET AND ENVIRONMENT INTERACTION

A. infant brain at birth



*FKBP5* - a key regulator of the hypothalamic-pituitary-adrenal (HPA) axis.

Question: Does *FKBP5* modulate the relationship between antenatal maternal depression and neonatal hippocampal volume?

Datasets: 161 mother-offspring dyads, brain images, and genotype data.

## SNPs in *FKBP5*

SNP	<i>p</i> -value
rs10807151	0.011
<b>rs3800373</b>	<b>0.002</b>
rs7757037	0.006
<b>rs9296158</b>	<b>0.001</b>
rs3777747	0.012
rs6926133	0.006
rs9380524	0.914
<b>rs4713904</b>	<b>0.002</b>
<b>rs9470080</b>	<b>7.27×10<sup>-4</sup></b>
<b>rs9380526</b>	<b>5.48×10<sup>-4</sup></b>
rs10456432	0.151
<b>rs9380529</b>	<b>0.002</b>
rs9394314	0.017
<b>rs2766533</b>	<b>3.94×10<sup>-4</sup></b>
rs12200498	0.029
rs2817032	0.008
rs7751693	0.281



# A MIXED EFFECT MODEL FOR GENE SET AND ENVIRONMENT INTERACTION (MIXGE)

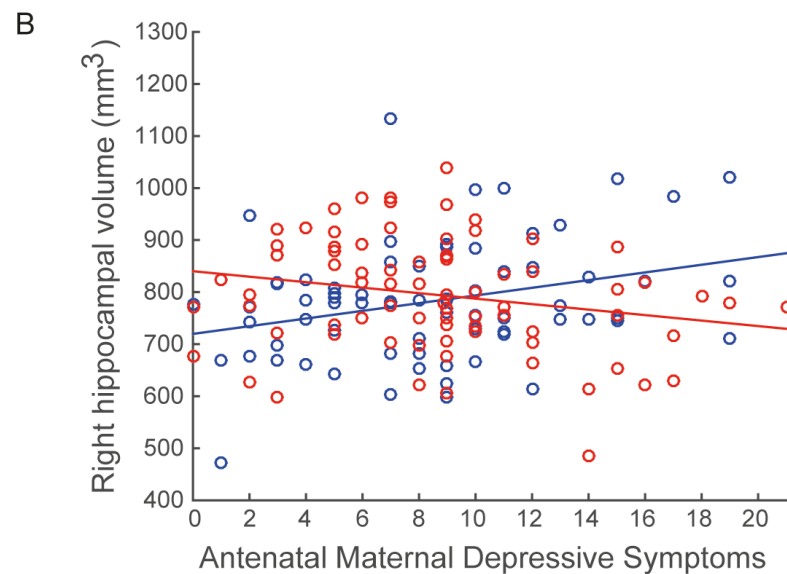
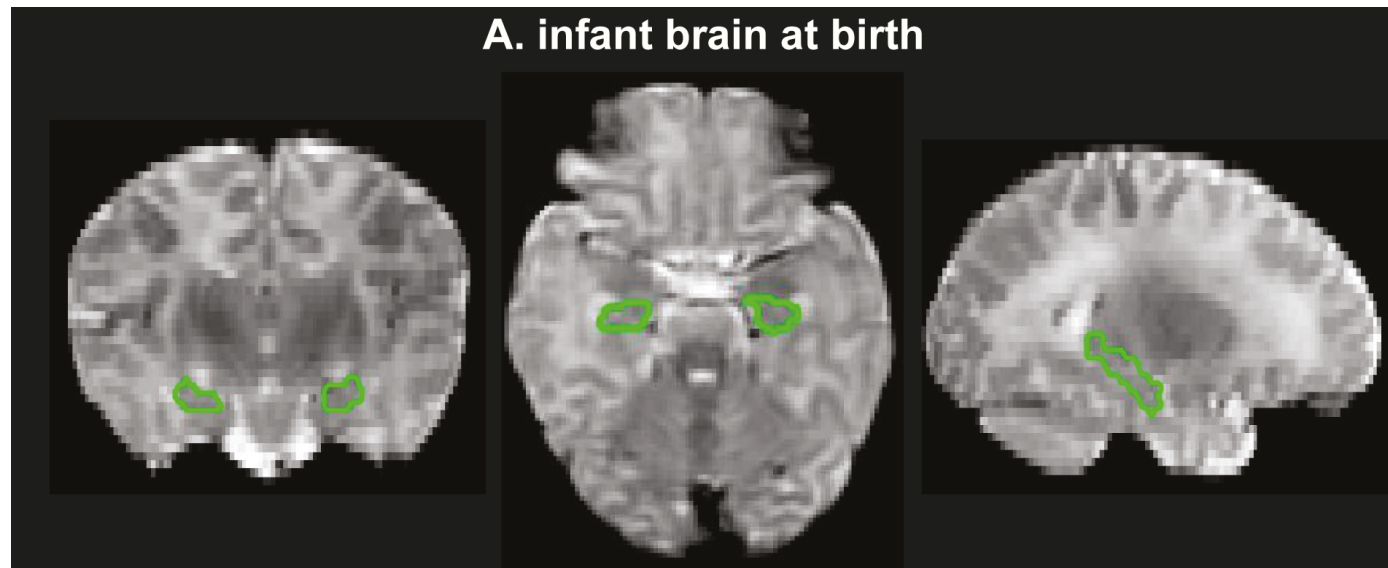
$$g\{E(y_i)\} = X_i^T \beta_x + e_i \beta_e + G_i^T W \pi^{(1)} + e_i G_i^T W \pi^{(2)} + e_i G_i^T \delta$$

Diagram illustrating the components of the mixed effect model for gene set and environment interaction (MIXGE):

- $X_i^T \beta_x$ : covariates
- $e_i \beta_e$ : maternal depression
- $G_i^T W \pi^{(1)}$ : gene set
- $e_i G_i^T W \pi^{(2)}$ : genetic score
- $e_i G_i^T \delta$ : interaction term



# FKBP5 MODULATES THE ASSOCIATION BETWEEN ANTENATAL MATERNAL DEPRESSIVE SYMPTOMS AND NEONATAL HIPPOCAMPAL VOLUME

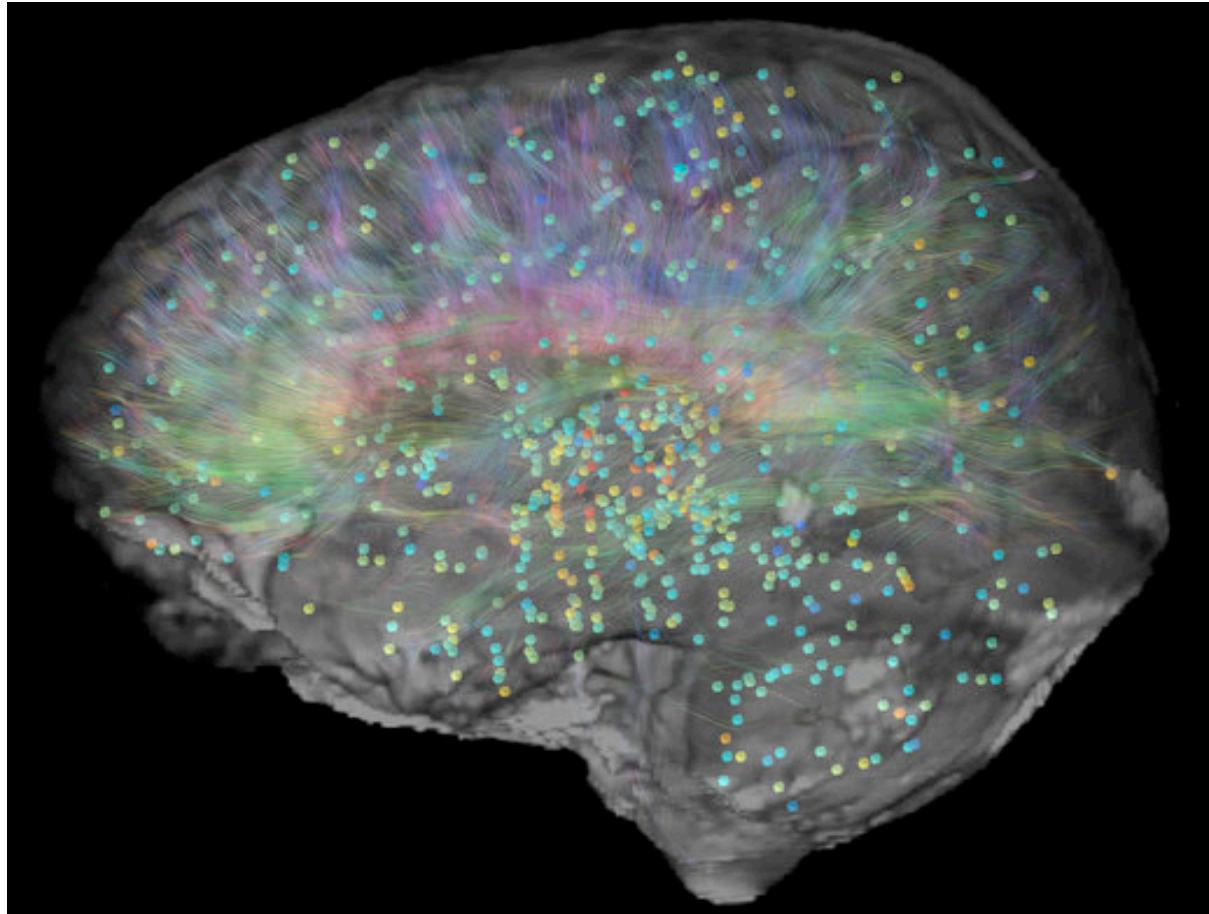




# GENE IS EXPRESSED IN SPECIFIC BRAIN REGIONS AT A SPECIFIC TIME

---

spatial pattern of gene expression



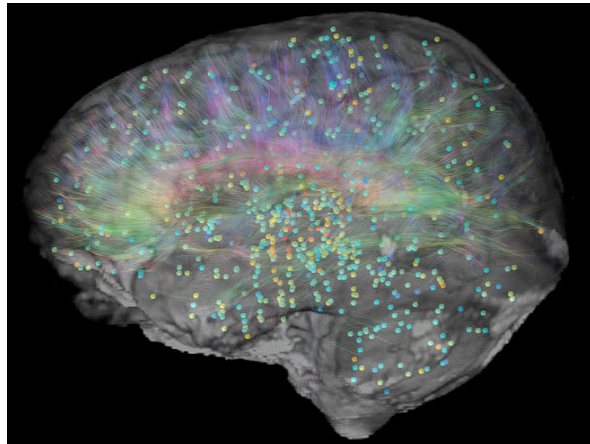


**NEURAL TRANSCRIPTION CORRELATES  
WITH CORTICAL MORPHOLOGY,  
REFLECTING THE BIOLOGICAL PROCESSES  
RELATED TO BRAIN DEVELOPMENT**

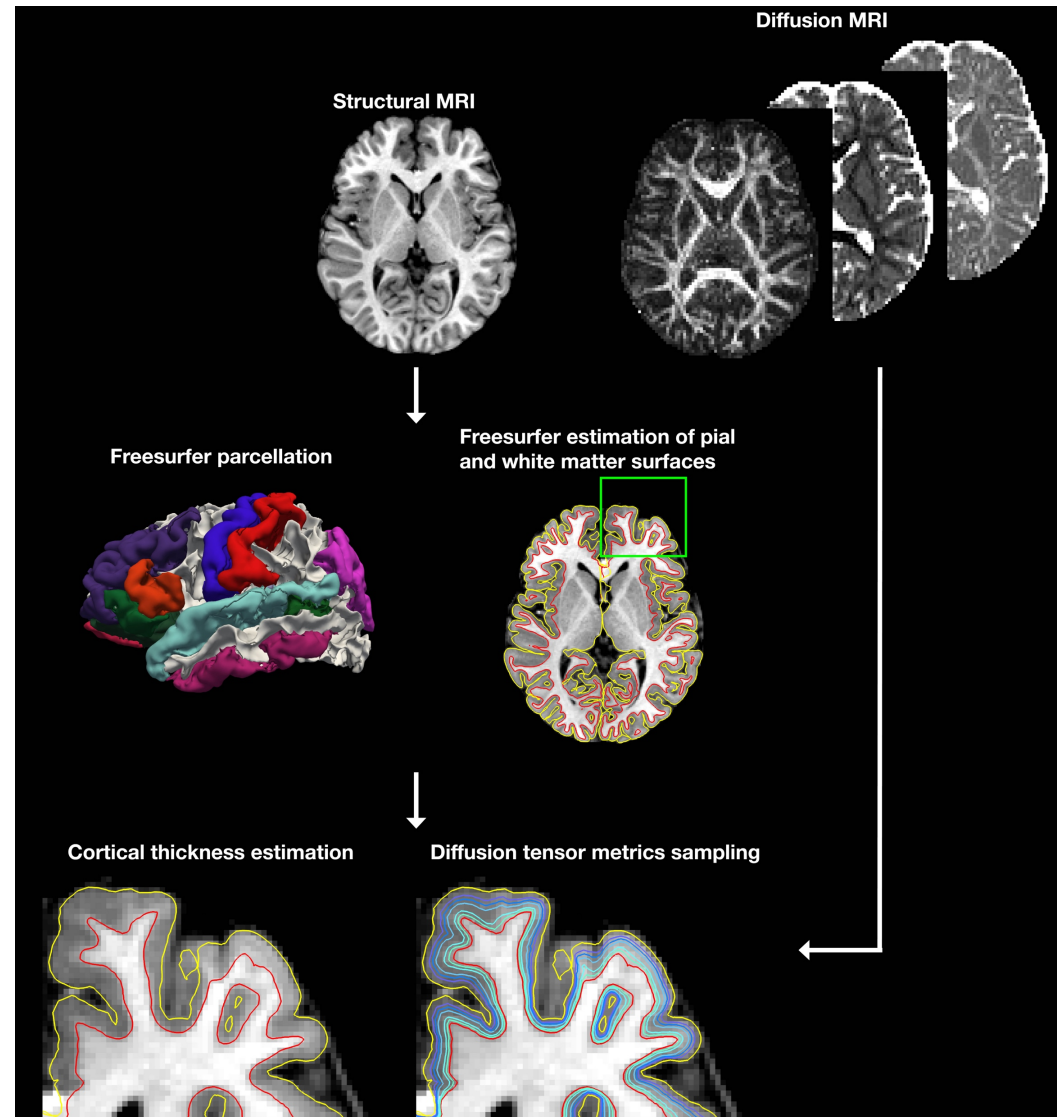


# NEURAL TRANSCRIPTION CORRELATES OF MULTIMODAL CORTICAL PHENOTYPES DURING DEVELOPMENT

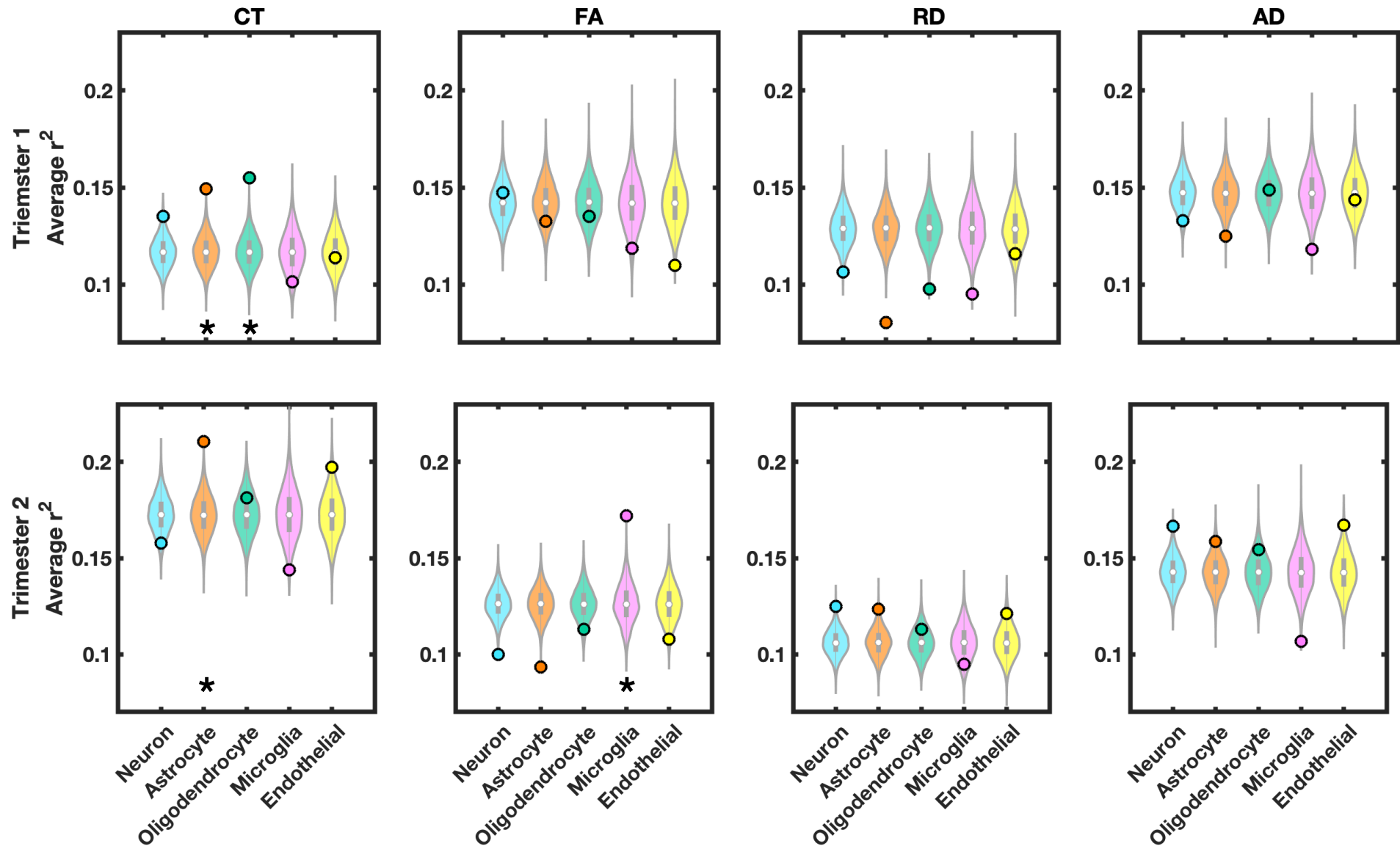
gene expression



multi-modal MRI data

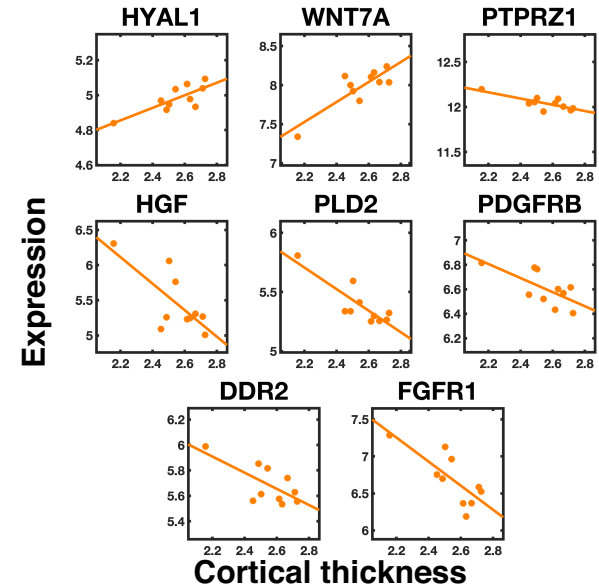
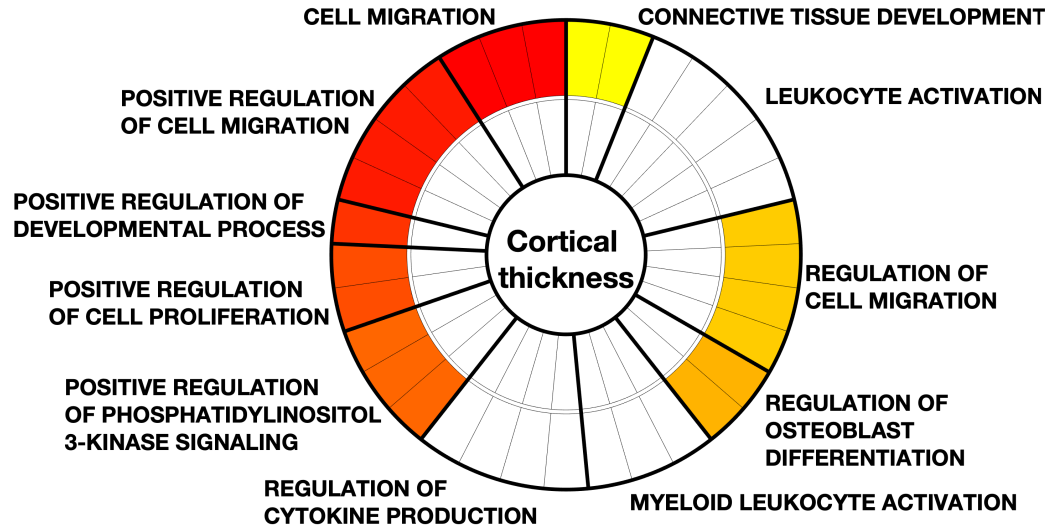


# NEONATAL IMAGING PHENOTYPES ARE ASSOCIATED WITH GLIAL GENE EXPRESSION

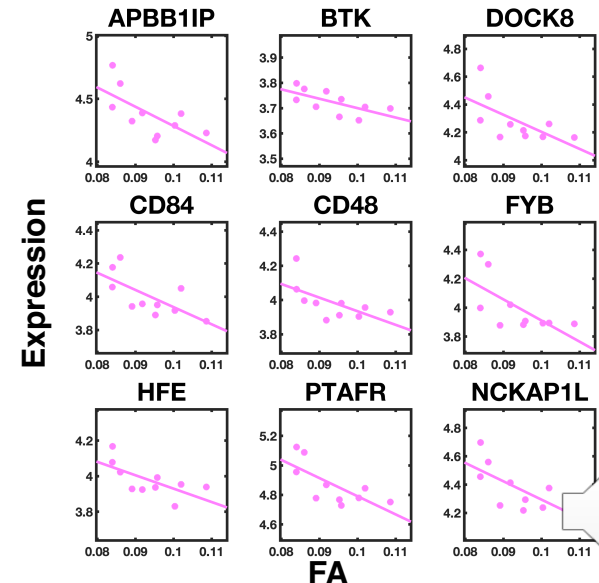
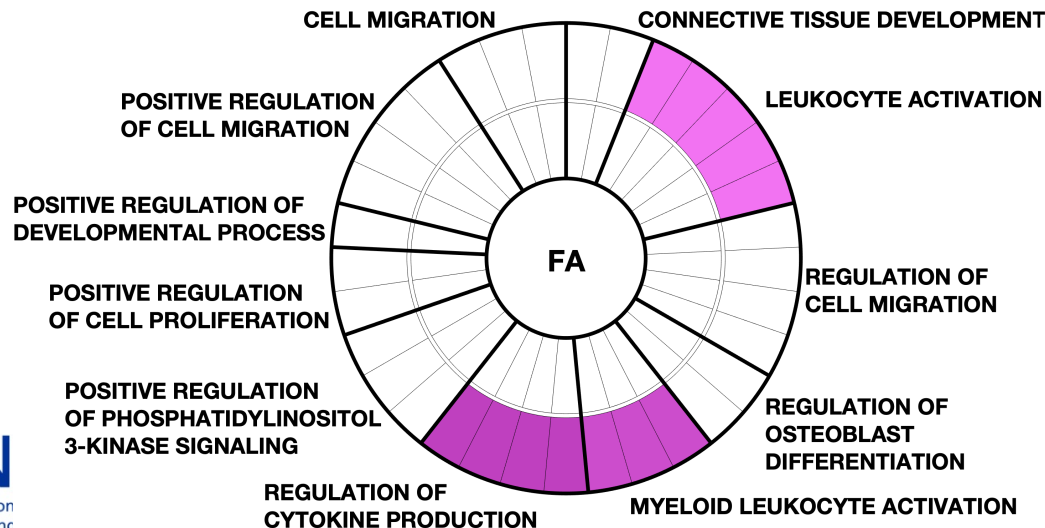


# NEONATAL IMAGING PHENOTYPES ARE ASSOCIATED WITH GLIAL GENE EXPRESSION

**Astrocyte**



**Microglia**



# BIOLOGICAL PROCESSES IN BRAIN DEVELOPMENT

## CELL TYPE

Neuron	Neuron
Astrocyte	Astrocyte
Oligodendrocyte	Oligodendrocyte
Microglia	Microglia
Endothelial	Endothelial

1<sup>st</sup> trimester

2<sup>nd</sup> trimester

3<sup>rd</sup> trimester

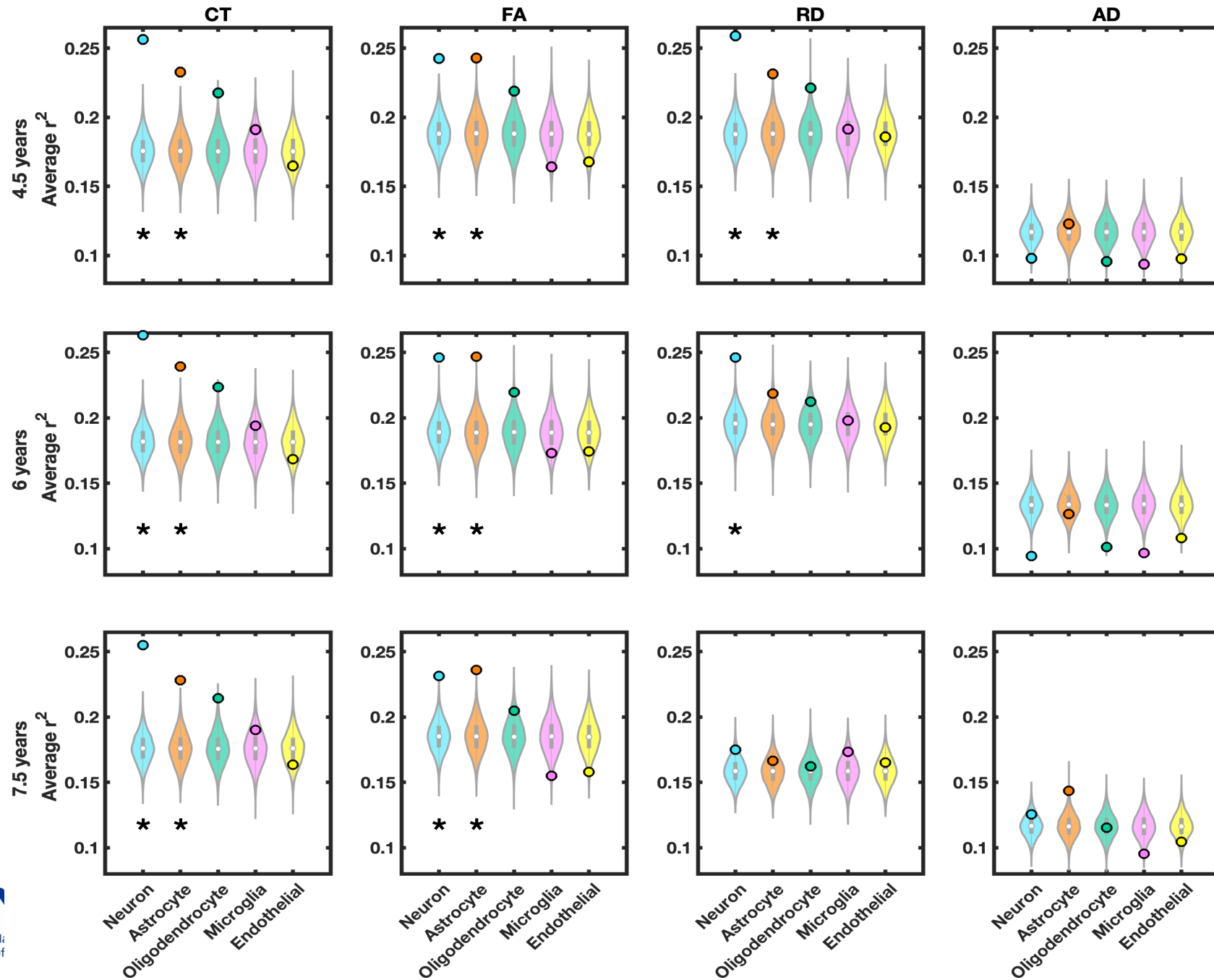
early childhood

## BIOLOGICAL PROCESSES

cell migration  
cell proliferation  
kinase signaling  
vascular development

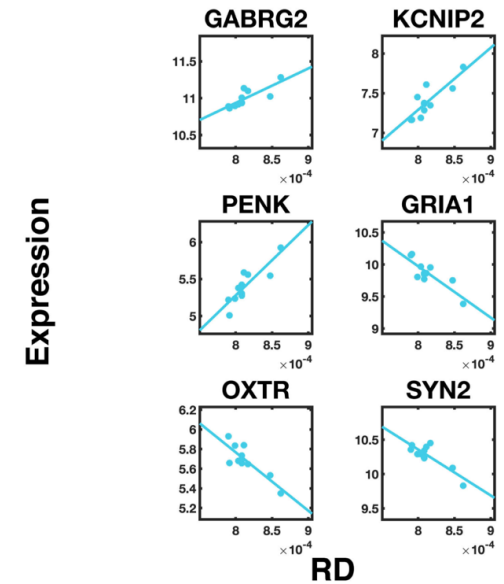
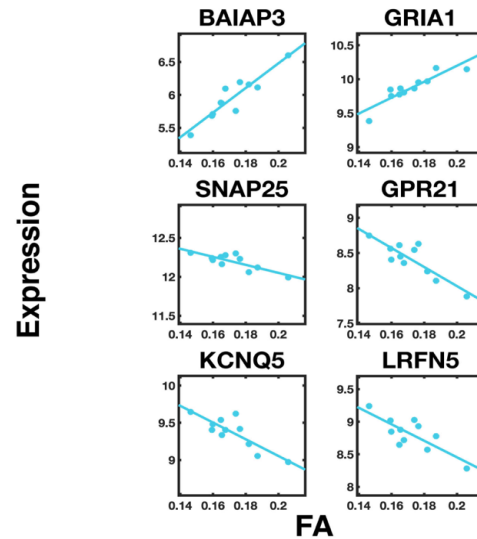
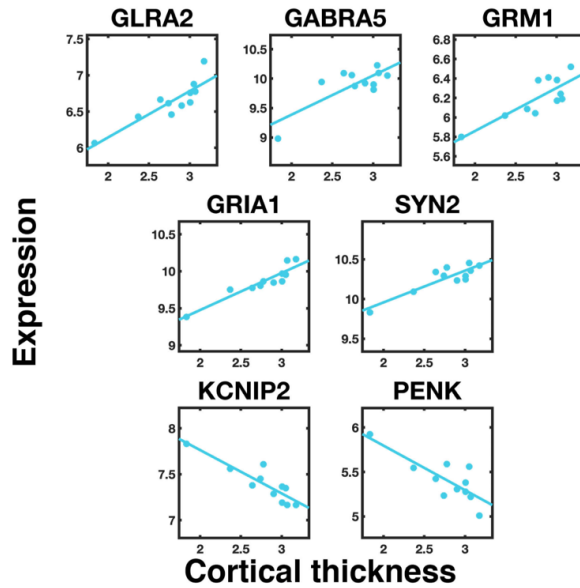
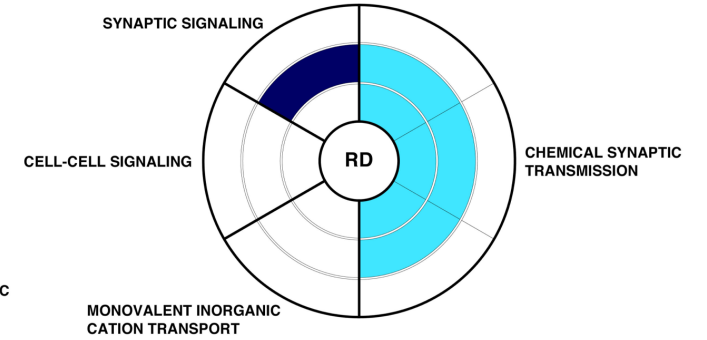
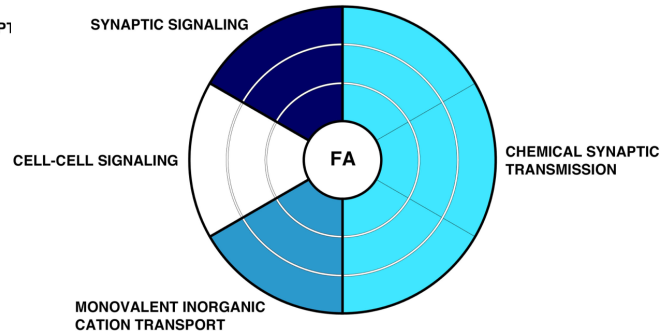
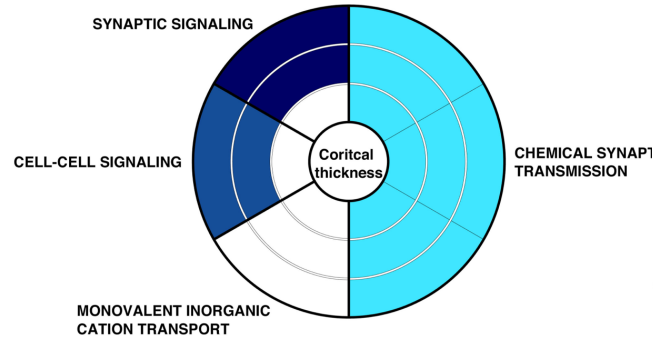


# CHILDHOOD IMAGING PHENOTYPES ARE ASSOCIATED WITH NEURON AND ASTROCYTE GENE SET EXPRESSION

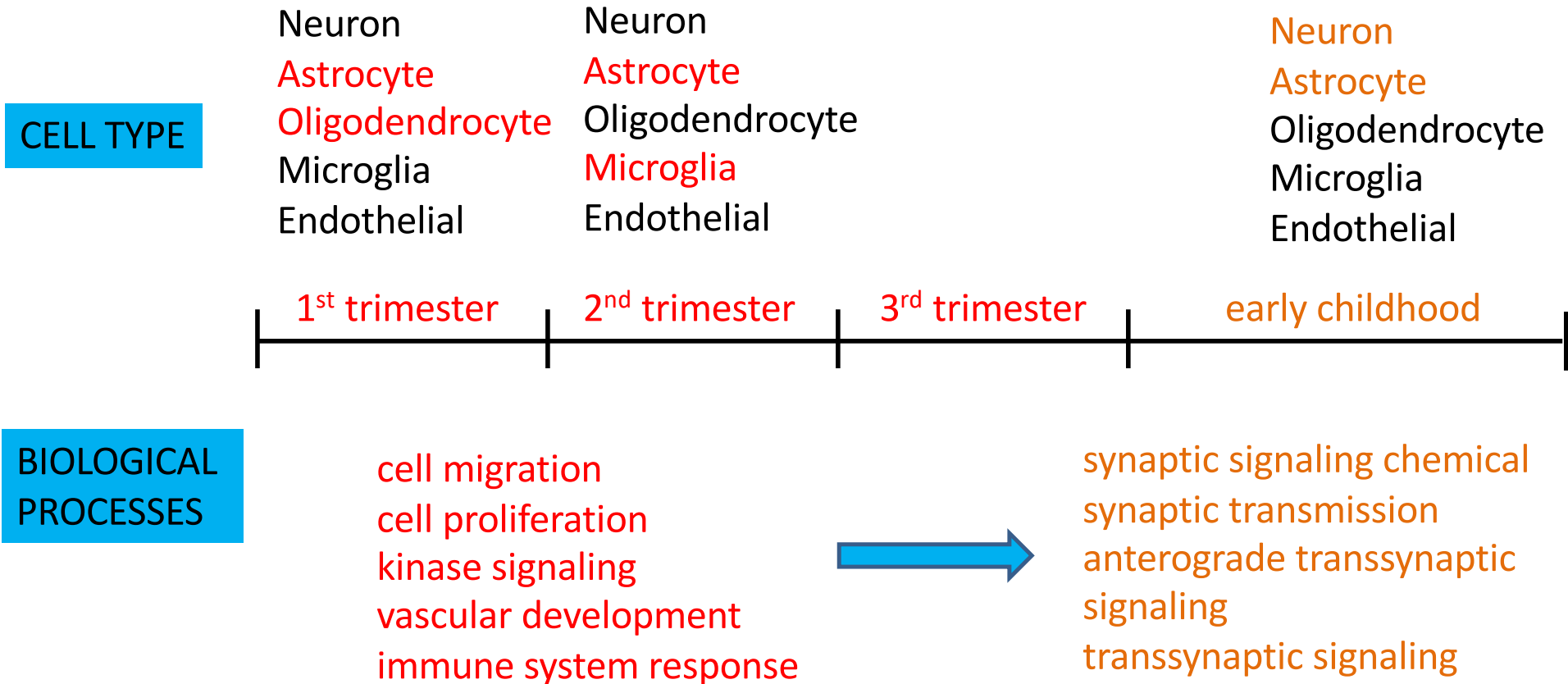


# CHILDHOOD IMAGING PHENOTYPES ARE ASSOCIATED WITH NEURON AND ASTROCYTE GENE SET EXPRESSION

Neuron

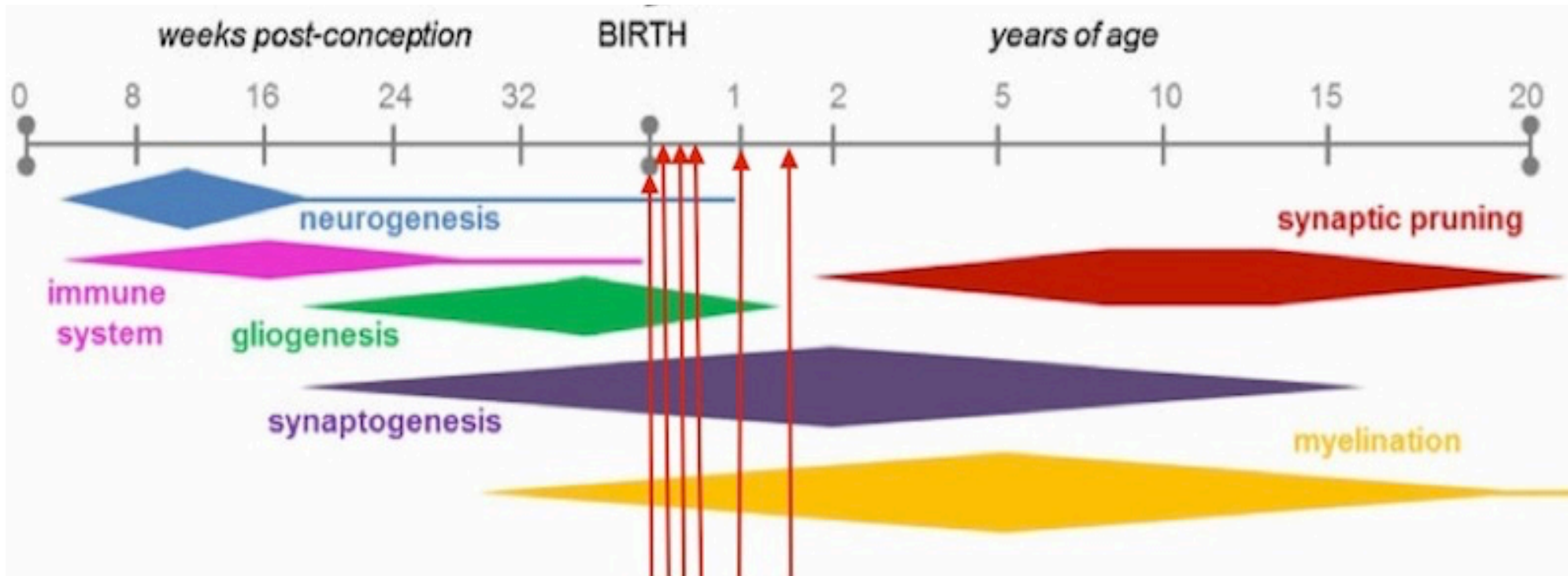


# BIOLOGICAL PROCESSES IN BRAIN DEVELOPMENT





# BIOLOGICAL PROCESSES IN BRAIN DEVELOPMENT



# POPULATION NEUROIMAGING MEETS GENETICS

---

- IMPORTANT CONSIDERATIONS – CONFOUNDING, SMALL SAMPLE SIZE, RARE VARIANTS
- POSSIBLE SOLUTIONS WITH DATA, BOTH BIG AND SMALL
  - a) CONFOUNDING ESTIMATION
  - b) POLYGENIC RISK SCORE
  - c) HAPLOTYPE
  - d) GENE-SET ANALYSIS
  - e) GENE EXPRESSION



# ACKNOWLEDGEMENTS

All my collaborators in ADNI and Tian Ge.

Laboratory for Medical Image Data Sciences at  
the National University of Singapore

