# Genome-wide analysis of DNA methylation in samples from the Genotype-Tissue Expression (GTEx) project

Peter Hickey

@PeteHaitch

~~Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health~~

Single Cell Open Research Endeavour (SCORE),

Walter and Eliza Hall Institute of Medical Research

**Slides: [www.bit.ly/AGTA2018](www.bit.ly/AGTA2018)**

# GTEx to eGTEx via a ‘pilot’ study

*The Genotype-Tissue Expression (GTEx) project is an ongoing effort to build a comprehensive public resource to study [human] tissue-specific gene expression and regulation.*

- GTEx Consortium, 2015, *Science* **348,** 648–660

*[eGTEx] extends the GTEx project to combine gene expression with additional intermediate molecular measurements on the same tissues.*

- eGTEx Project, 2017, *Nat. Genet.* ***49***, *1664–1670*

*Hmm, this eGTEx study is gonna be huge.*

*And the human brain is hella cool.*

*Let's do a pilot study.*

- Artist's impression of conversation in Hansen and Feinberg labs, c. 2015

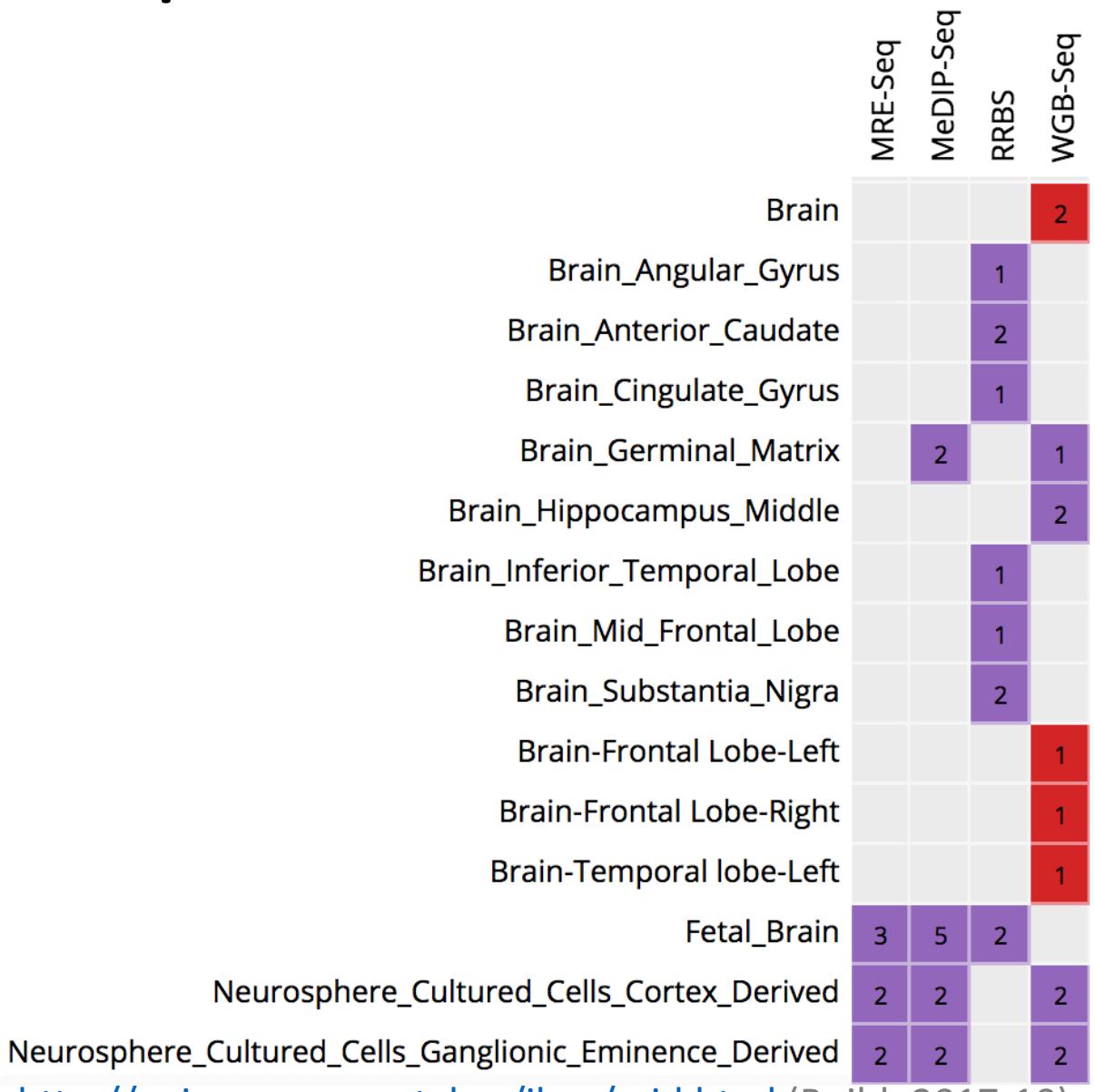
# BrainEpigenome (the ‘pilot’ study)

*Rizzardi, L*. Hickey, P.F.*, et al.* **Neuronal brain region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric disease heritability**.

*bioRxiv* (2017), doi:10.1101/120386 (in press, Nature Neuroscience)

UCSC Track Hub: www.bit.ly/BrainEpigenomeHub

# Map of human brain methylome was limited (c. 2015)




http://epigenomesportal.ca/ihec/grid.html (Build: 2017-10)

- Little whole genome bisulfite sequencing (WGBS) data
- Few (if any) biological replicates
- Mostly bulk tissue samples
- Few brain region-specific differentially methylated regions (DMRs)[1,2]

[1]Davies, M. N. *et al.* Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol.* **13,** R43 (2012).
[2]Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015).

# A good map requires biological replicates, multiple brain regions, and multiple cell types


**Tissue**

**BRNCTXB (frontal cortex)**

**BRNACC (anterior cingulate cortex)**

**BRNHPP (hippocampus)**

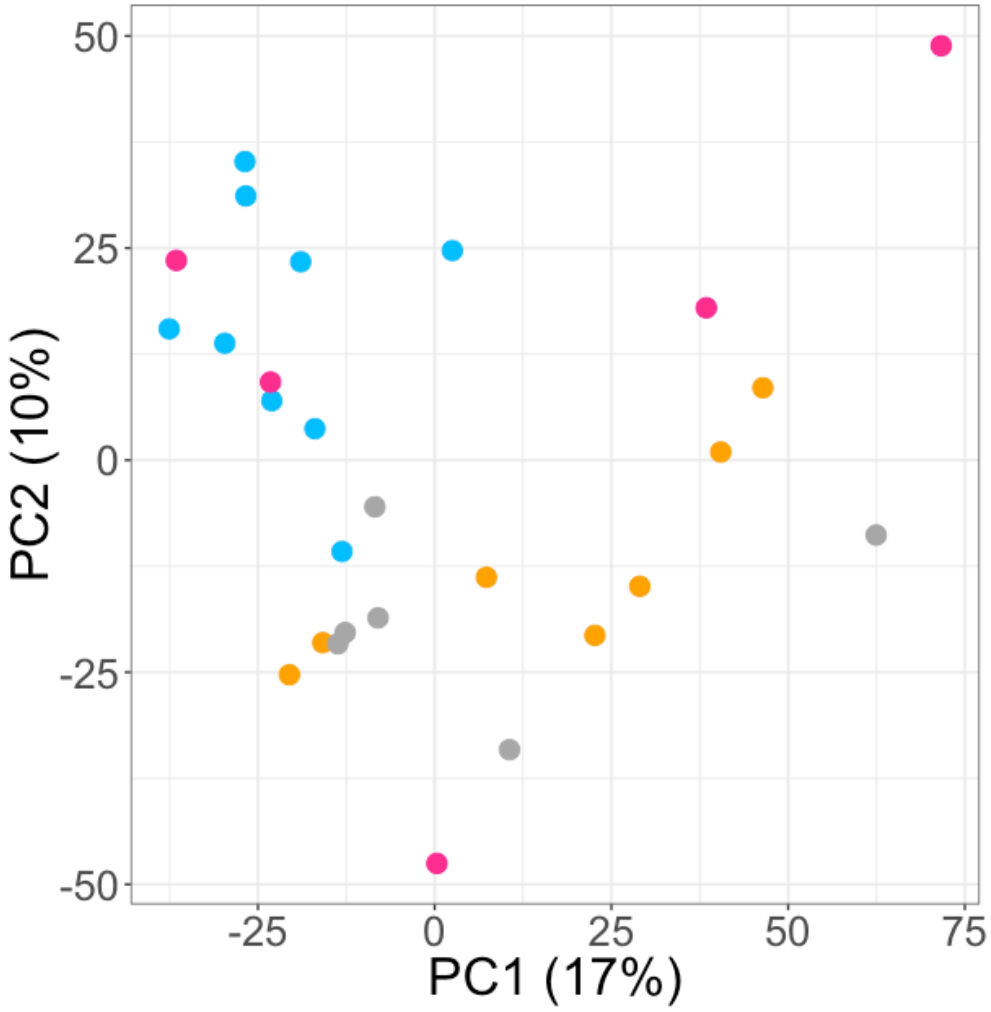
**BRNNCC (nucleus accumbens)**

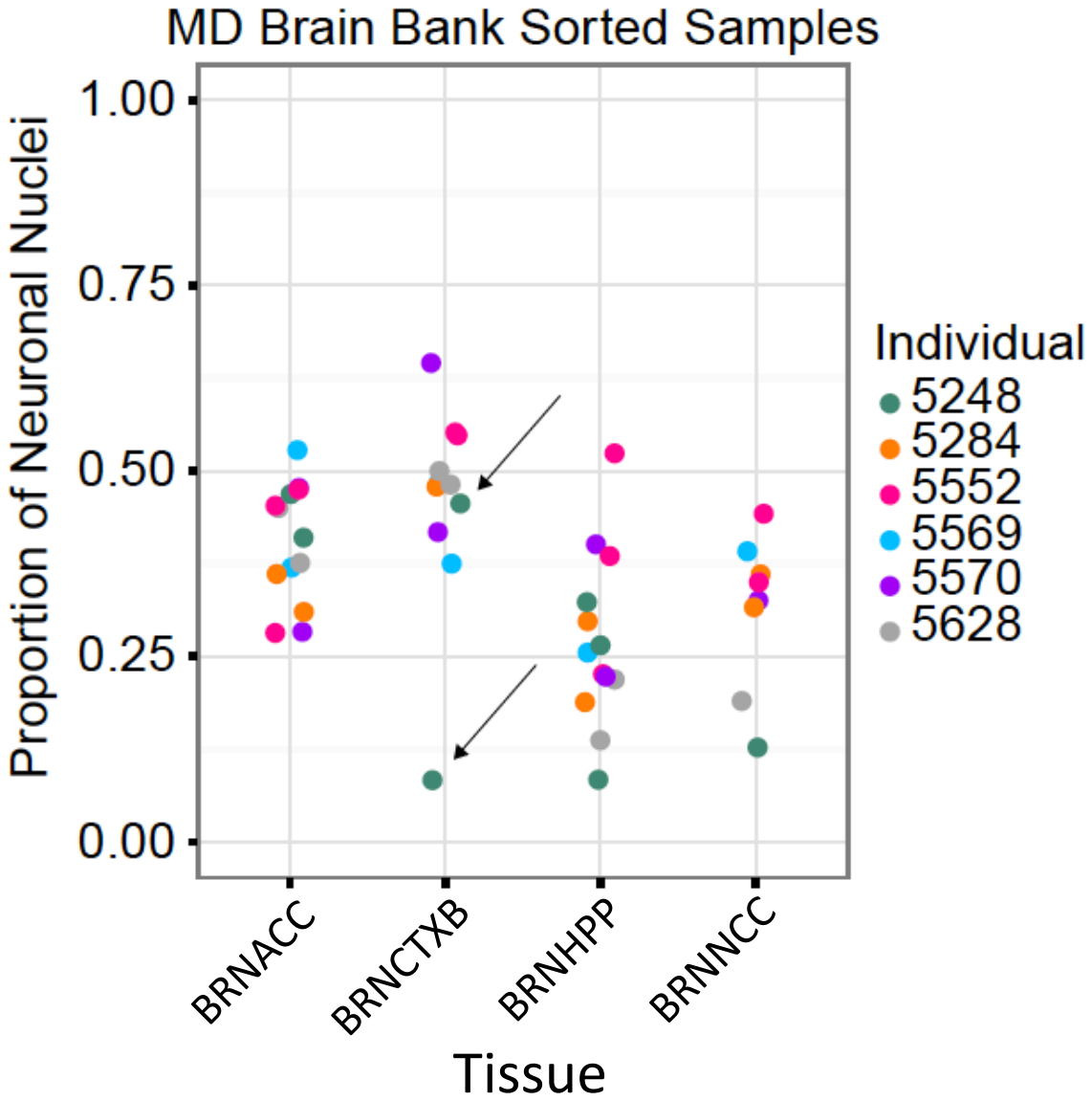
**Donor 1**

**Donor**

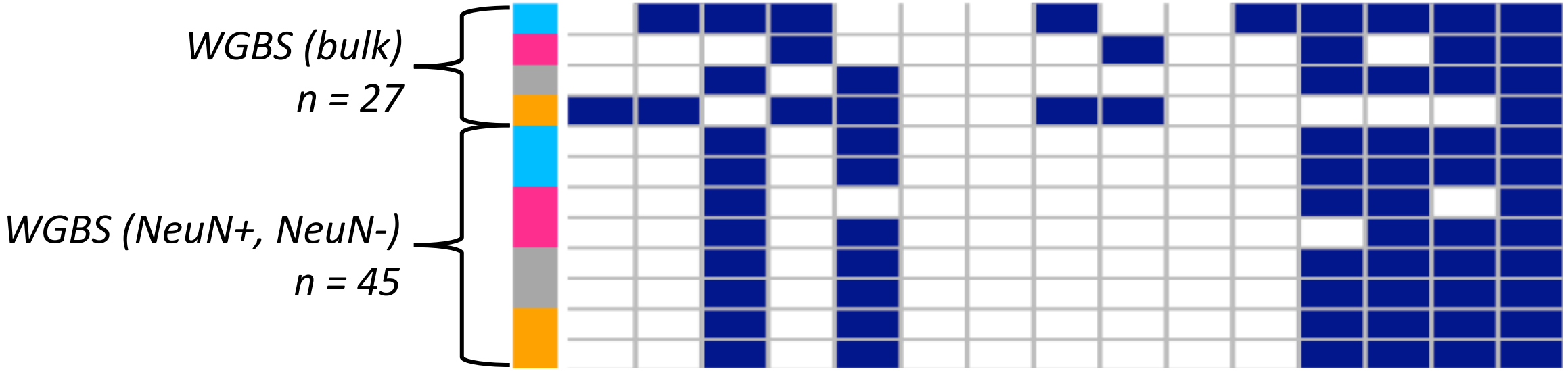
**Donor 15**

# Bulk tissue samples are uninformative for brain region-specific mCG due to variation of neuronal proportion in sampled tissue



# Let’s try fluorescence activated nuclei sorting (FANS)




**Tissue**

**BRNCTXB (frontal cortex)**

**BRNACC (anterior cingulate cortex)**

**BRNHPP (hippocampus)**

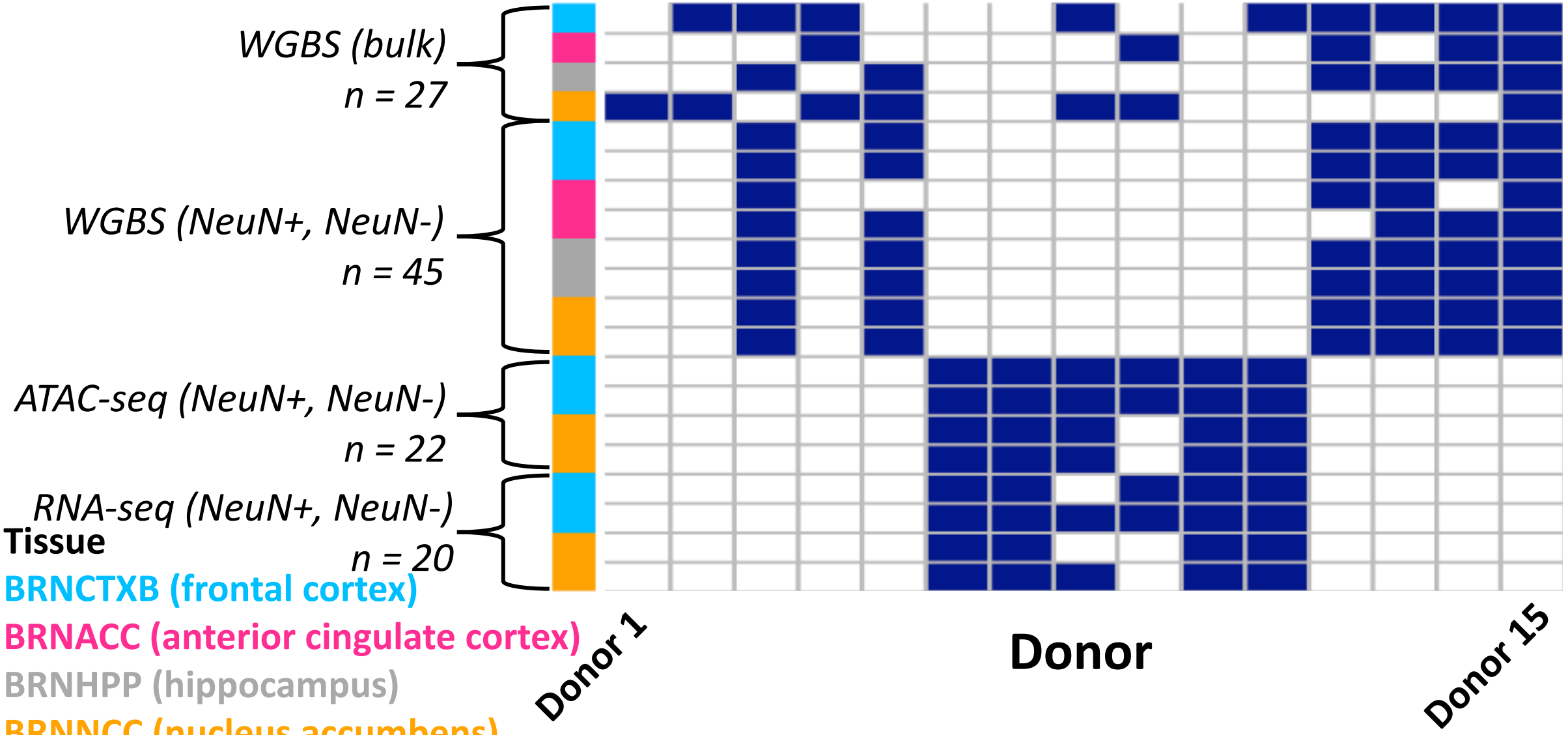
**BRNNCC (nucleus accumbens)**

**Donor 1**

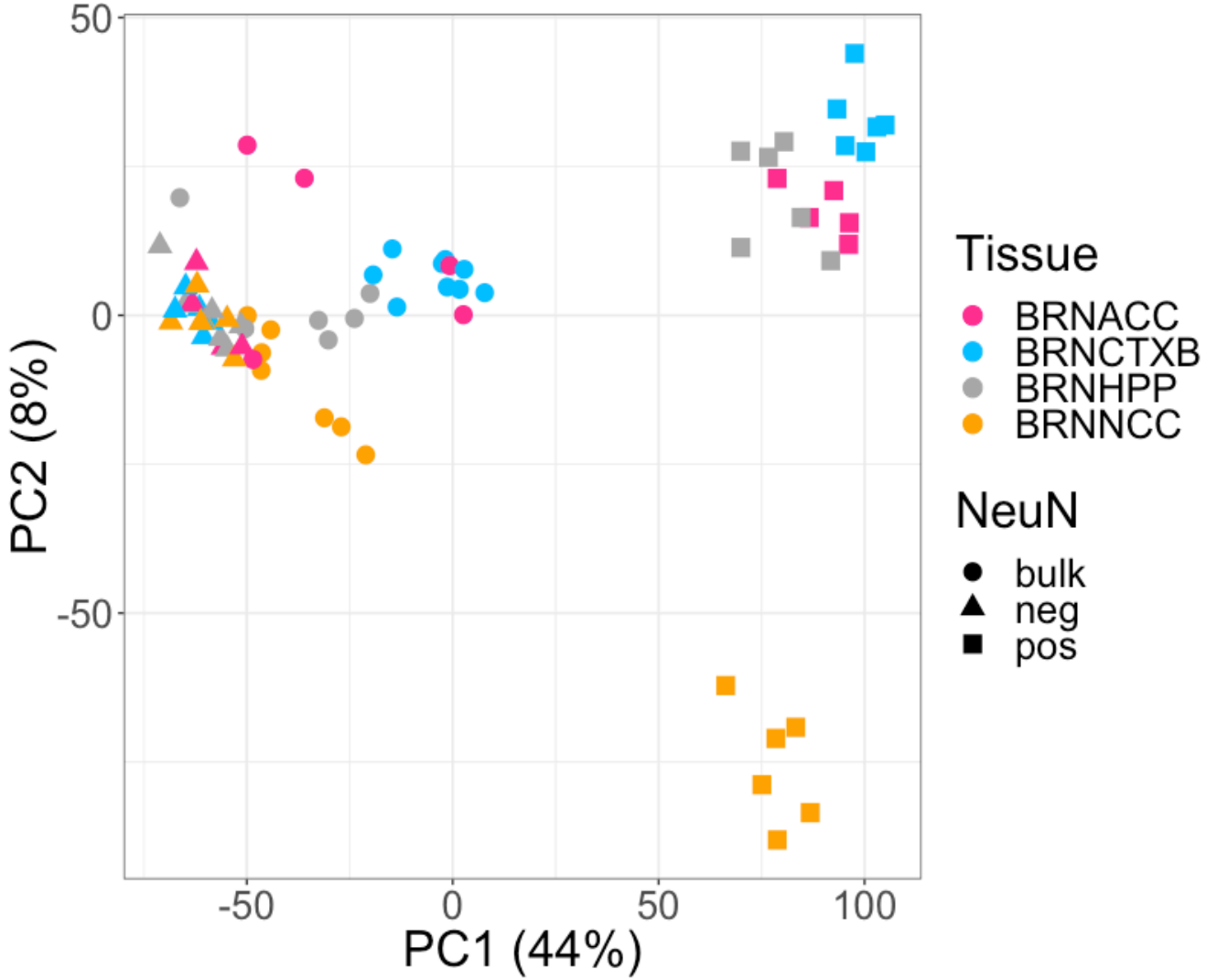
**Donor**

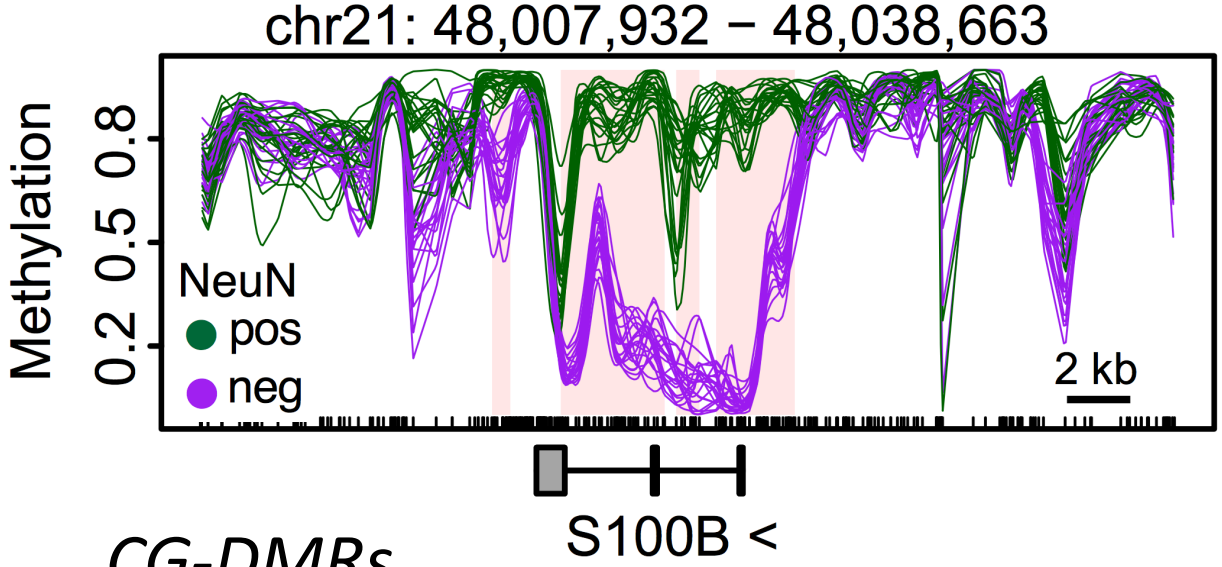
**Donor 15**

# And let’s do some more assays



# FANS & WGBS reveals brain region-specificity of mCG in NeuN+ (but not NeuN-) samples



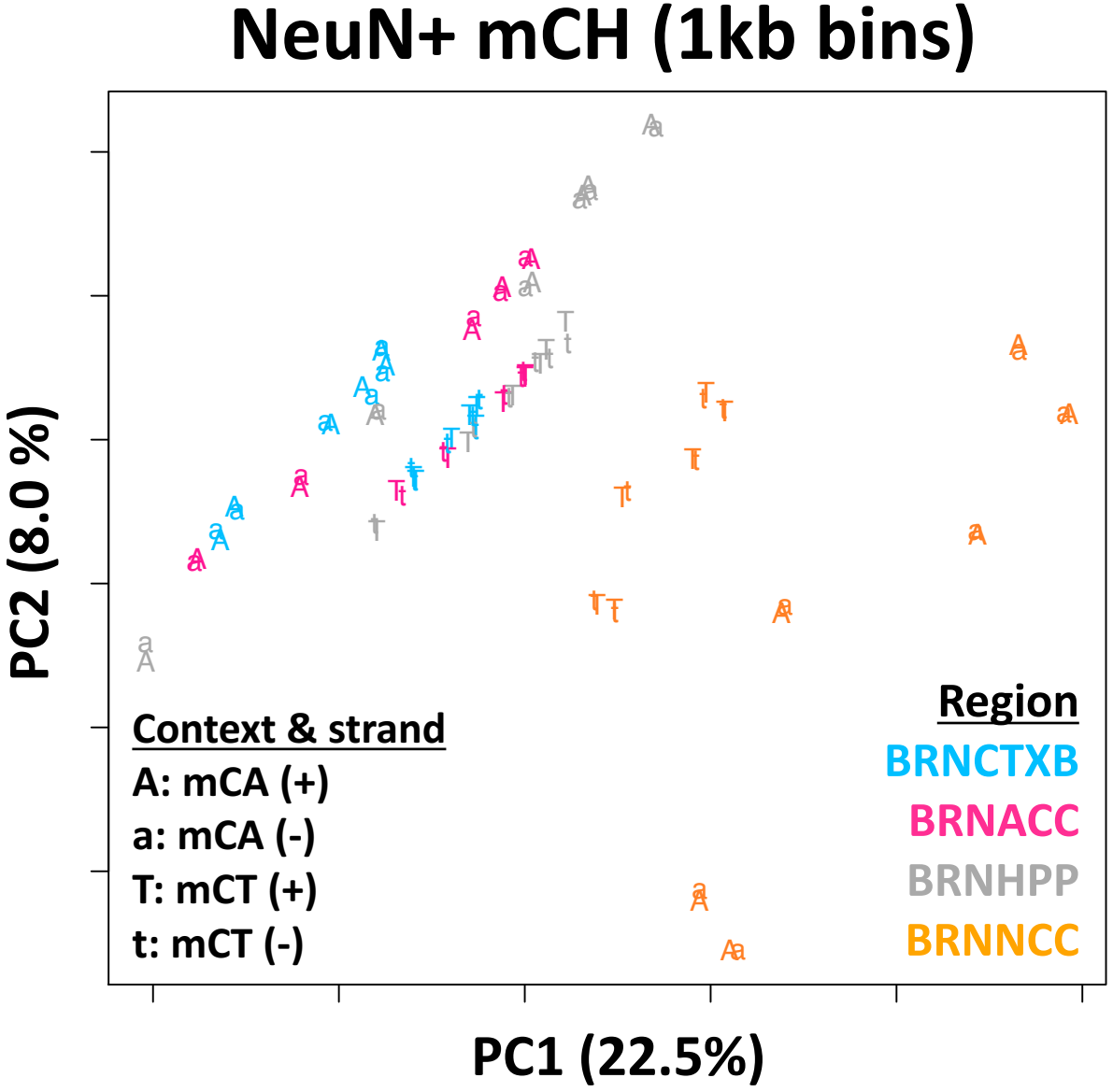



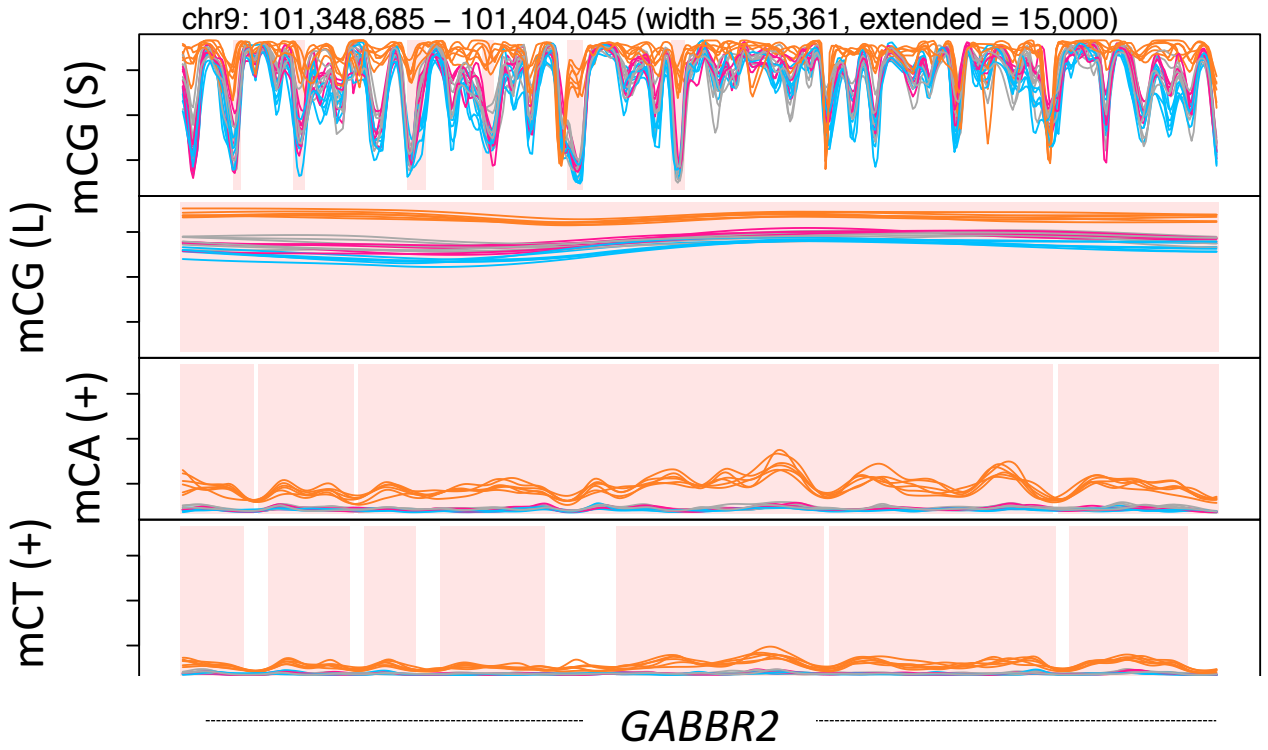

*CG-DMRs*

| | n | Size |
|---|---|---|
| NeuN+ vs. NeuN- | 100,875* | 70.0 Mb |
| NeuN+ | 13,074 | 11.9 Mb |
| NeuN- | 114 | 0.1 Mb |

*21,802 novel DMRs

# NeuN+ samples: mCH shows limited strand specificity, ‘tracks’ mCG, and can be used to identify CH-DMRs



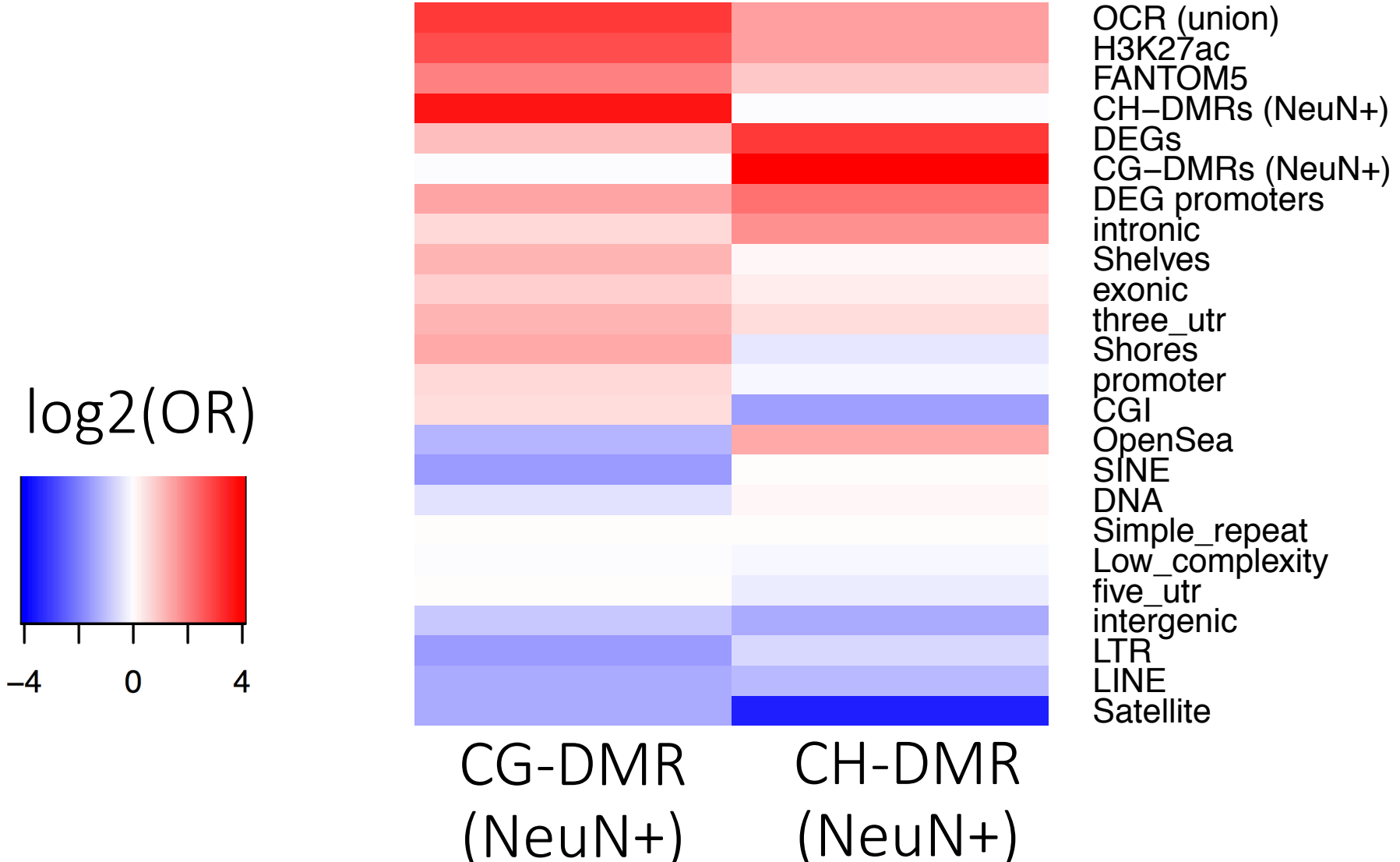

*CH-DMRs*

| | n | Size |
|---|---|---|
| NeuN+ | 15,029[+] | 39.6 Mb[++] |

[+]Before merging across strand and context

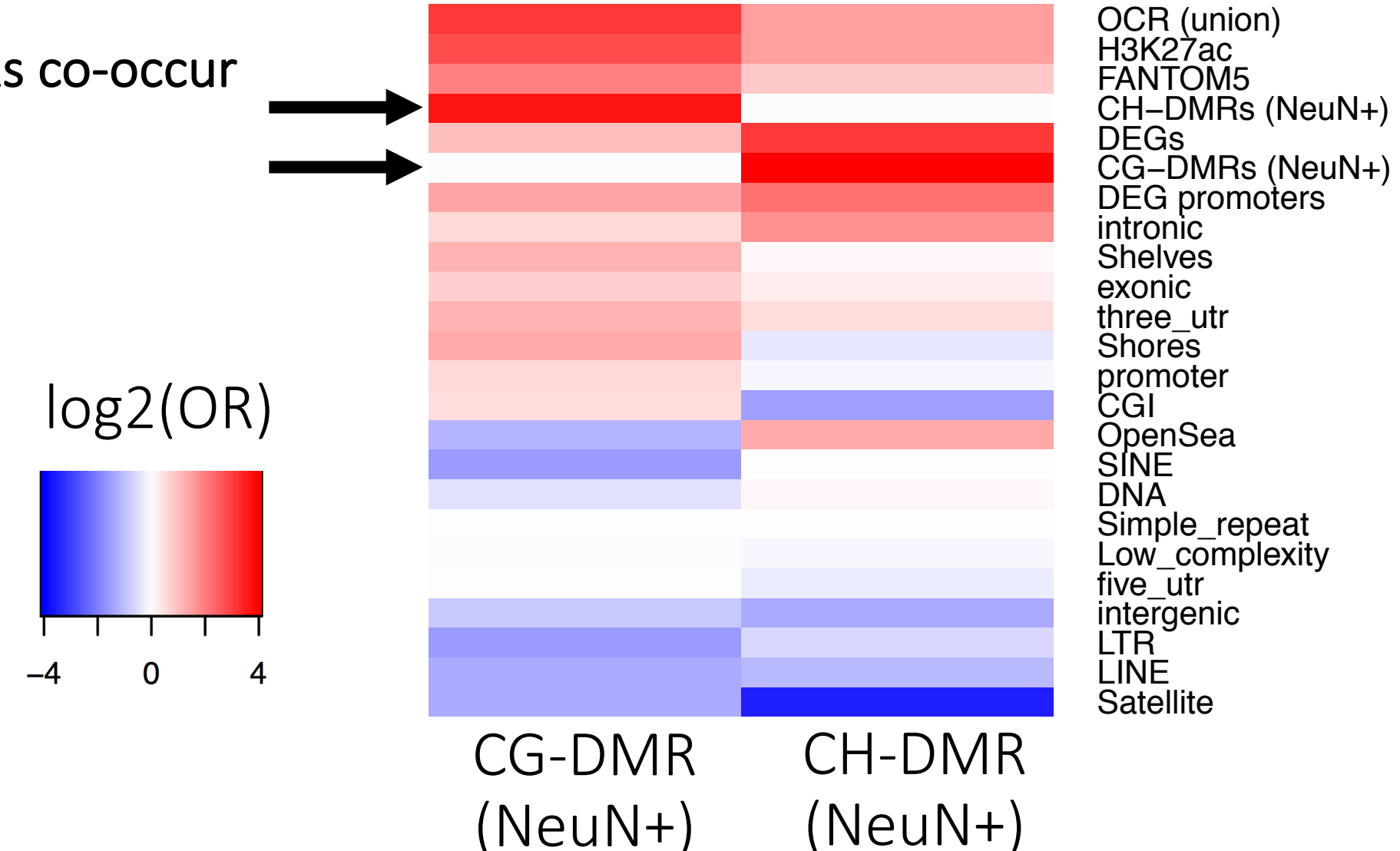
[++]After merging across strand and context

# Enrichment of DMRs over genomic features



# Enrichment of DMRs over genomic features

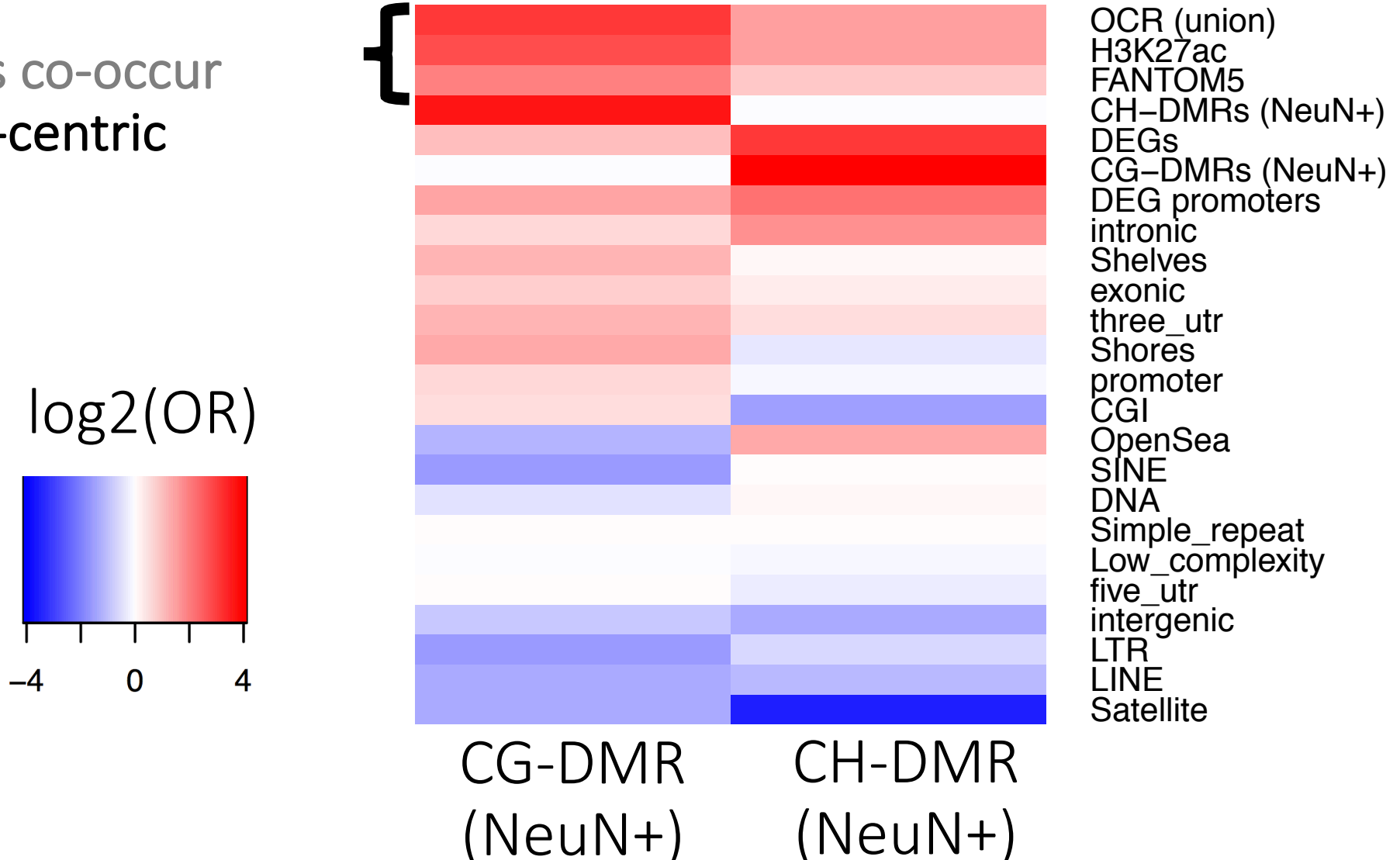
CG-DMRs and CH-DMRs co-occur



# Enrichment of DMRs over genomic features

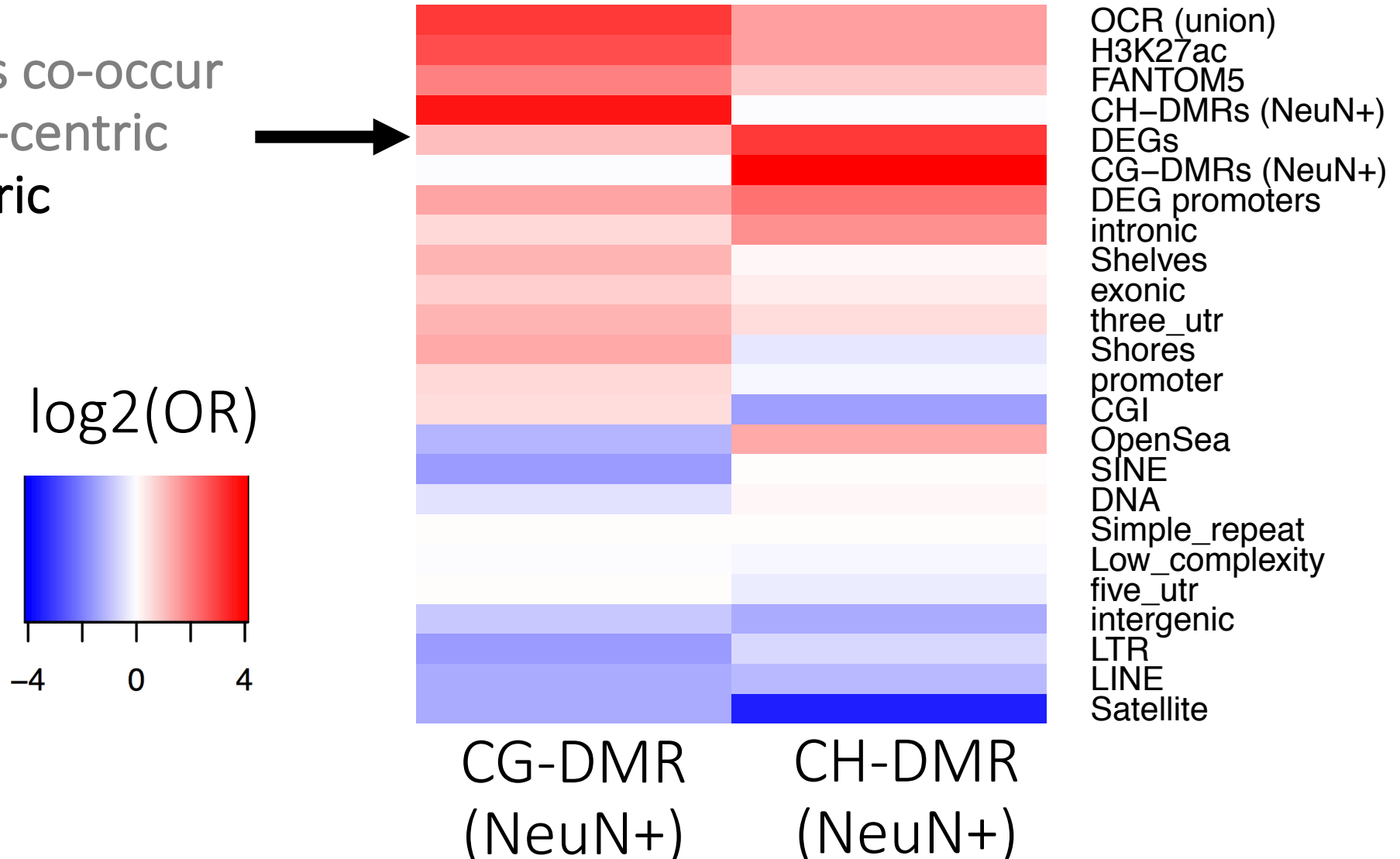
CG-DMRs and CH-DMRs co-occur

CG-DMRs are enhancer-centric



# Enrichment of DMRs over genomic features

CG-DMRs and CH-DMRs co-occur
CG-DMRs are enhancer-centric
CH-DMRs are DEG-centric

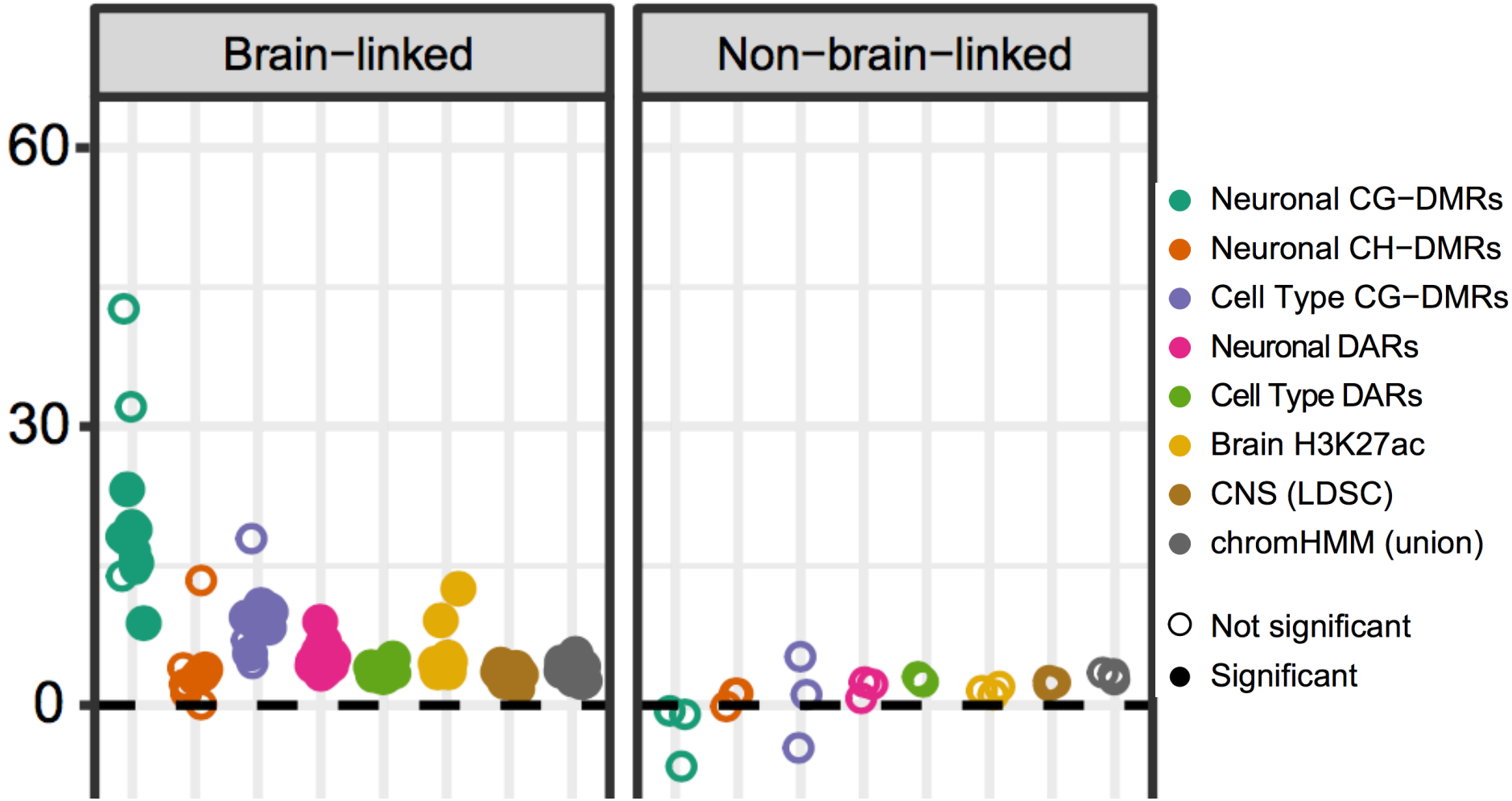


# CG-DMRs in NeuN+ samples are enriched for GWAS heritability of neuropsychiatric traits

Stratified linkage disequilibrium score regression*

27 ‘brain-linked’ traits (e.g., Schizophrenia, ADHD)

3 ‘non-brain-linked’ traits (e.g., height)




*Finucane, H. K. *et al.* **Partitioning heritability by functional annotation using genome-wide association summary statistics.** *Nat. Genet. (2015)* doi: 10.1038/ng.3404

# eGTEx (work in-progress)

*eGTEx Project* **Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease**.

*Nature Genetics (2017),* doi: 10.1038/ng.3969

# eGTEX study design

| Molecular phenotype | Primary assay(s) | Targeted tissues (phase II) | Targeted sample number |
|---|---|---|---|
| DNA accessibility | DNase I hypersensitivity | Brain regions, heart, lung, muscle, esophagus, breast, prostate, skin | ~1,135 |
| Histone modifications | ChIP–seq | Brain regions, heart, lung, muscle | ~600 |
| DNA methylation | WGBS and capture bisulfite sequencing | Brain regions, heart, lung, muscle, thyroid | ~2,000 |
| Allele-specific expression | mmPCR–seq | All tissues | ~2,000 |
| Post-transcriptional RNA modifications | $m^6A$ methylation capture sequencing | Brain regions, heart, lung, muscle | ~300 |
| Proteomic variation | MS, targeted arrays for transcription factors and cell signaling proteins | Brain, heart, lung, muscle, thyroid, colon, liver, prostate, pancreas, ovary, testis, breast | ~1,000 (MS)<br>~2,500 (arrays) |
| Somatic variation | Deep exome sequencing, RNA-seq, SNP arrays | ~20–25 tissues | ~800 |
| Telomere length | Luminex-based assay for telomere-repeat abundance | ~20 tissues | ~5,000 |

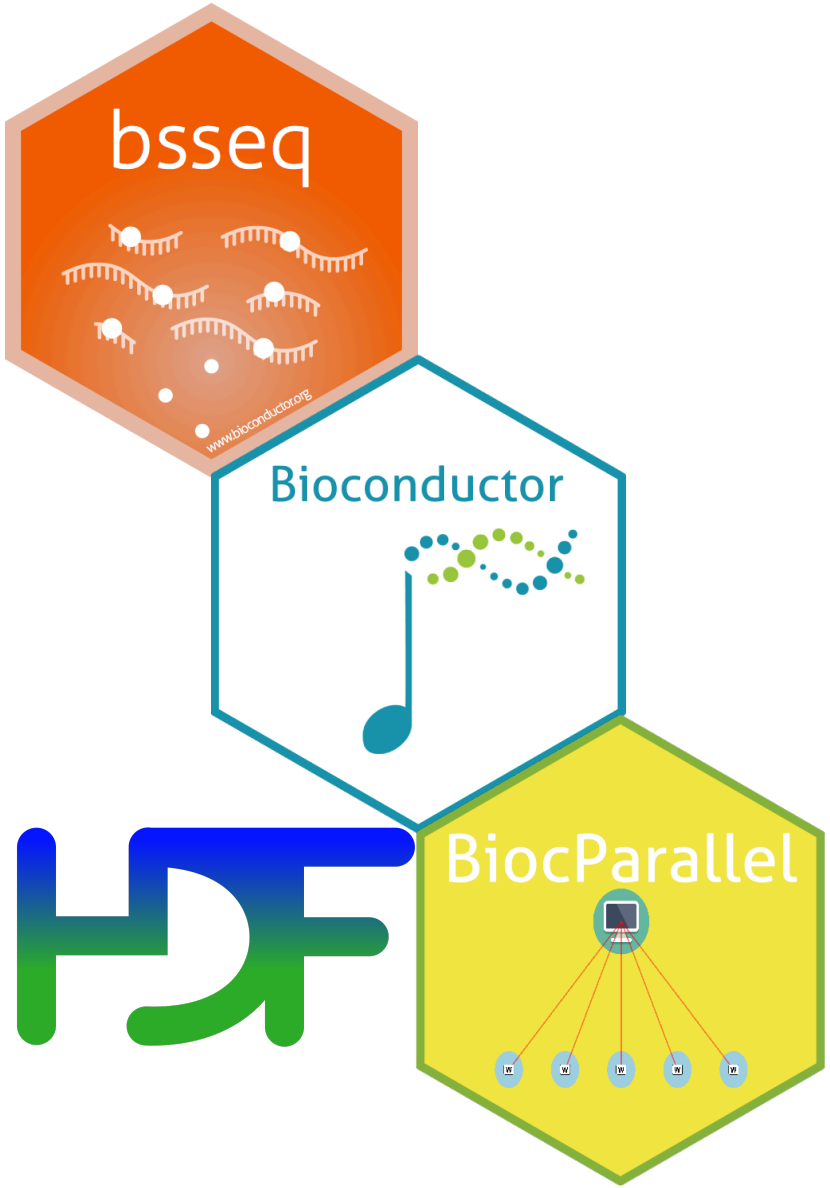
Molecular assays, targeted tissues, and sample number for eGTEx.

*eGTEx Project* **Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease.**

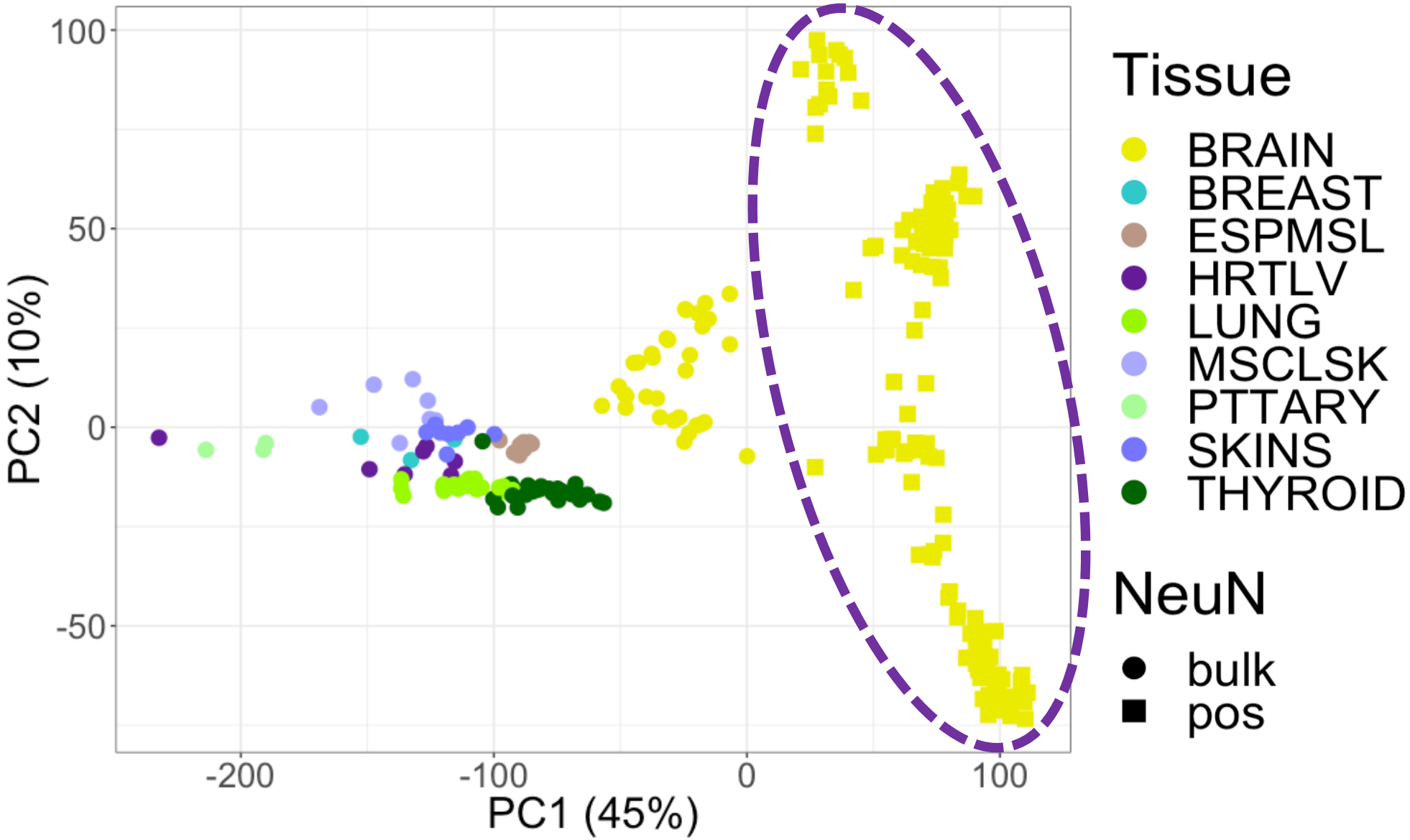
*Nature Genetics (2017),* doi: 10.1038/ng.3969

# Re-wrote *bsseq* to process and analyse eGTEx-sized (and bigger) datasets

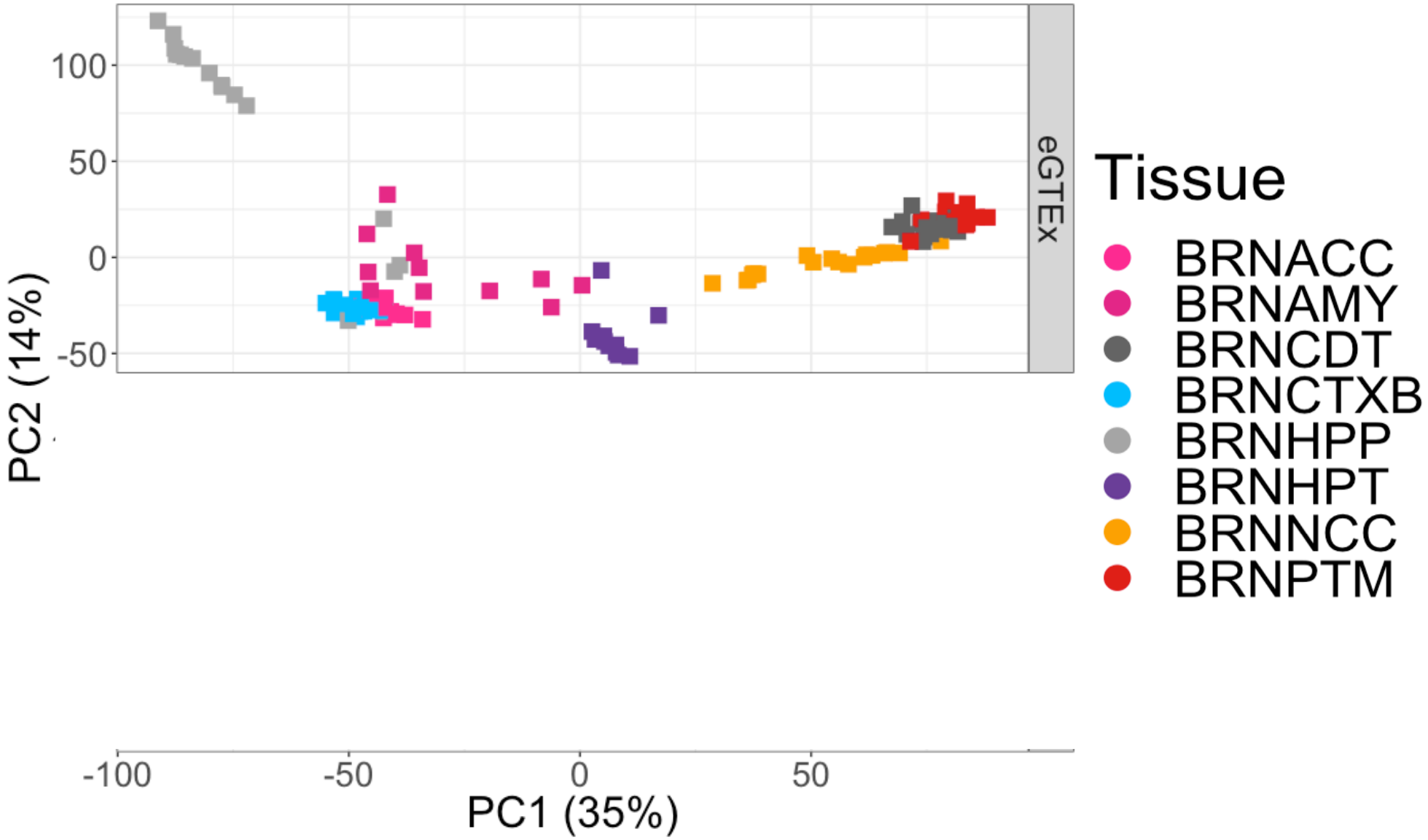
- Processed data is too large to store and operate on in-memory (10s – 100s of GB)
  - Data stored on-disk in HDF5 file
- Improved parallelization of key steps
  - Importing files
  - Smoothing
  - DMR calling
  - Permutation testing
- Available through Bioconductor
  - https://bioconductor.org/packages/bsseq/



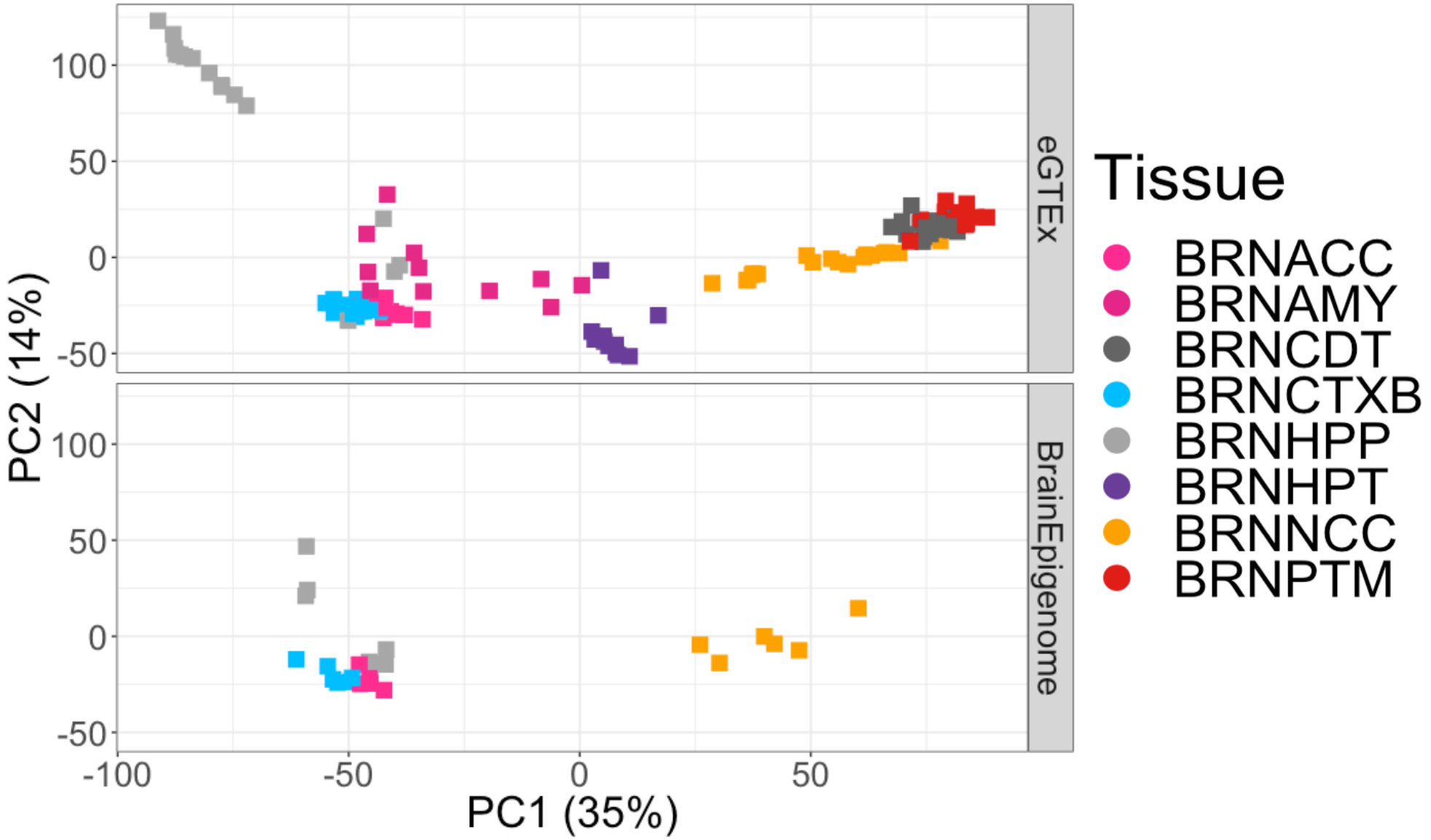
# mCG distinguishes eGTEx samples by tissue



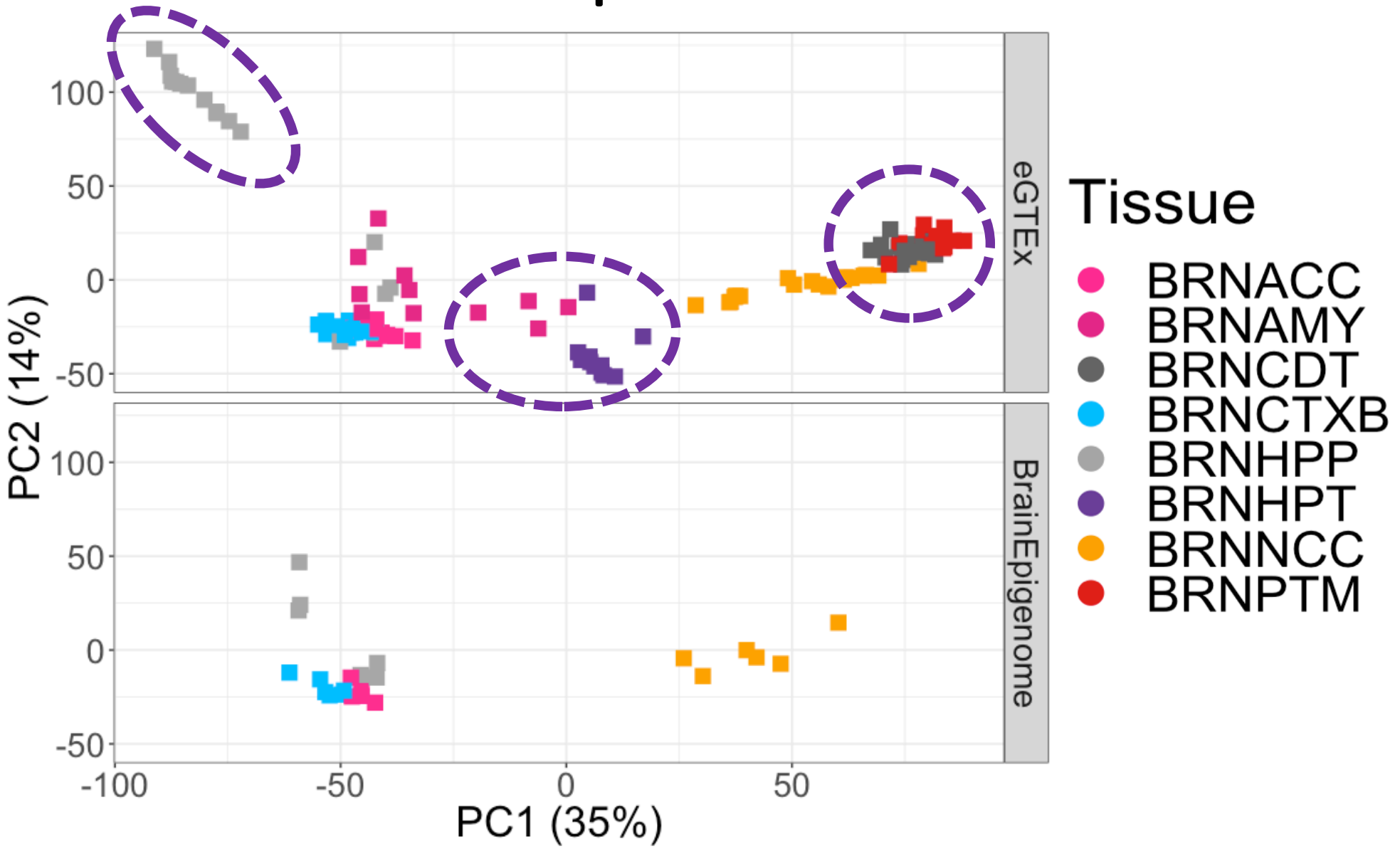
# eGTEx NeuN+ samples are (mostly) consistent with BrainEpigenome NeuN+ samples



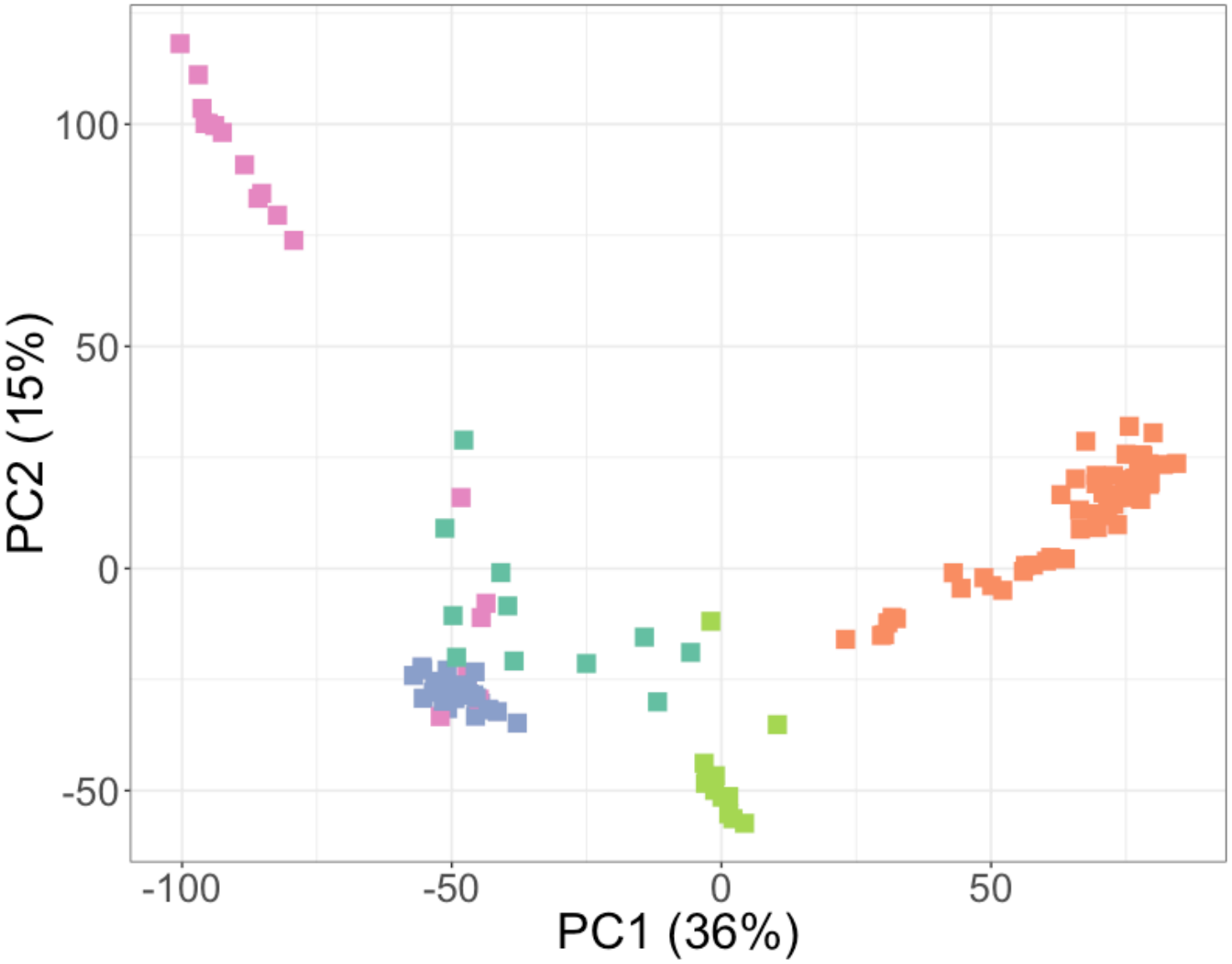
# eGTEx NeuN+ samples are (mostly) consistent with BrainEpigenome NeuN+ samples



# eGTEx NeuN+ samples are (mostly) consistent with BrainEpigenome NeuN+ samples



# 5-group: 16x as many CG-DMRs in eGTEx NeuN+ samples as in BrainEpigenome NeuN+ samples

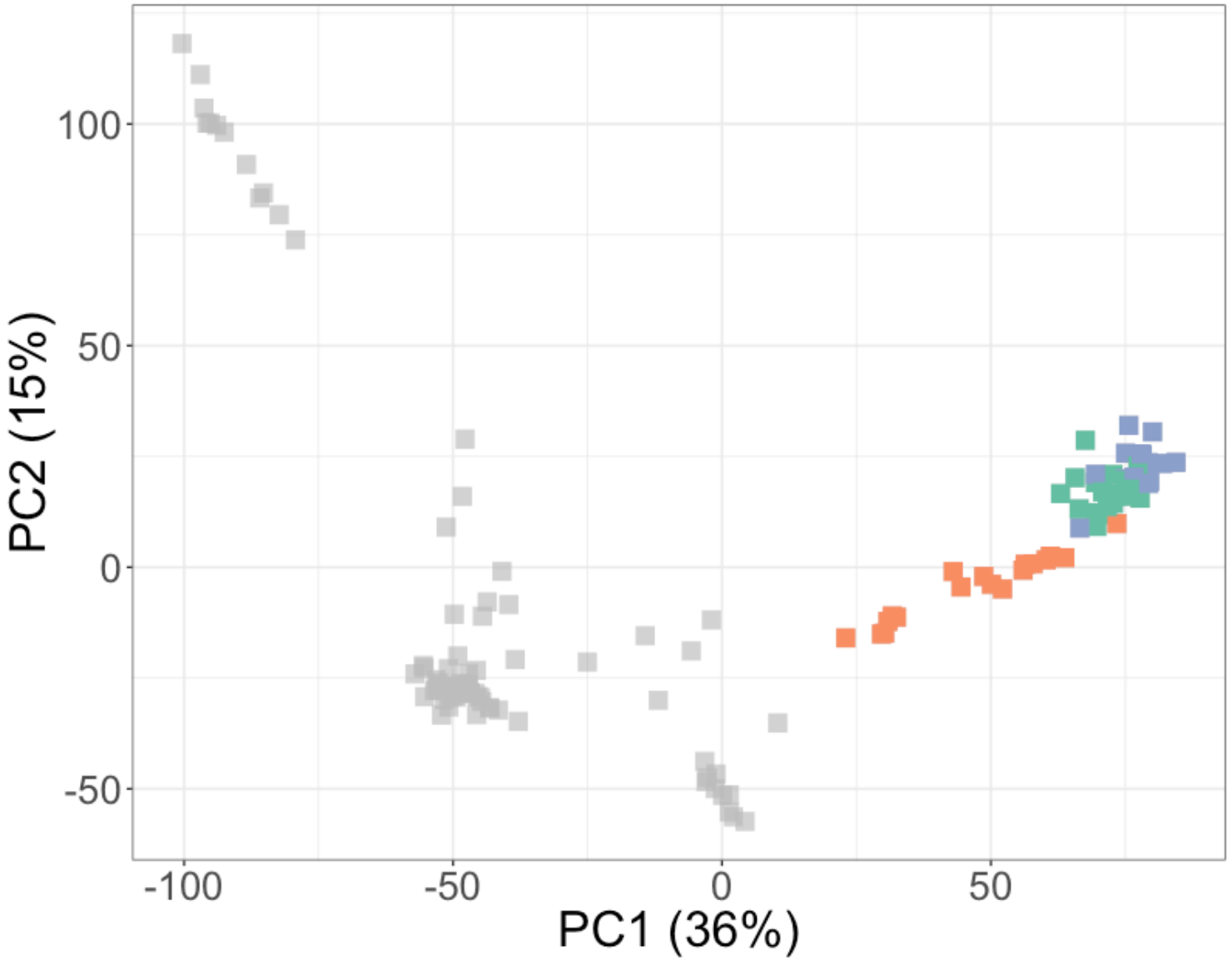



Group
- Amygdala
- Basal ganglia
- Cortical
- Hippocampus
- Hypothalamus

*CG-DMRs*

| | n | Size |
|---|---|---|
| 5-group | 181,146 | 196.9 Mb |

# Basal ganglia: Discover 2x as many CG-DMRs in eGTEx NeuN+ samples as in BrainEpigenome NeuN+ samples

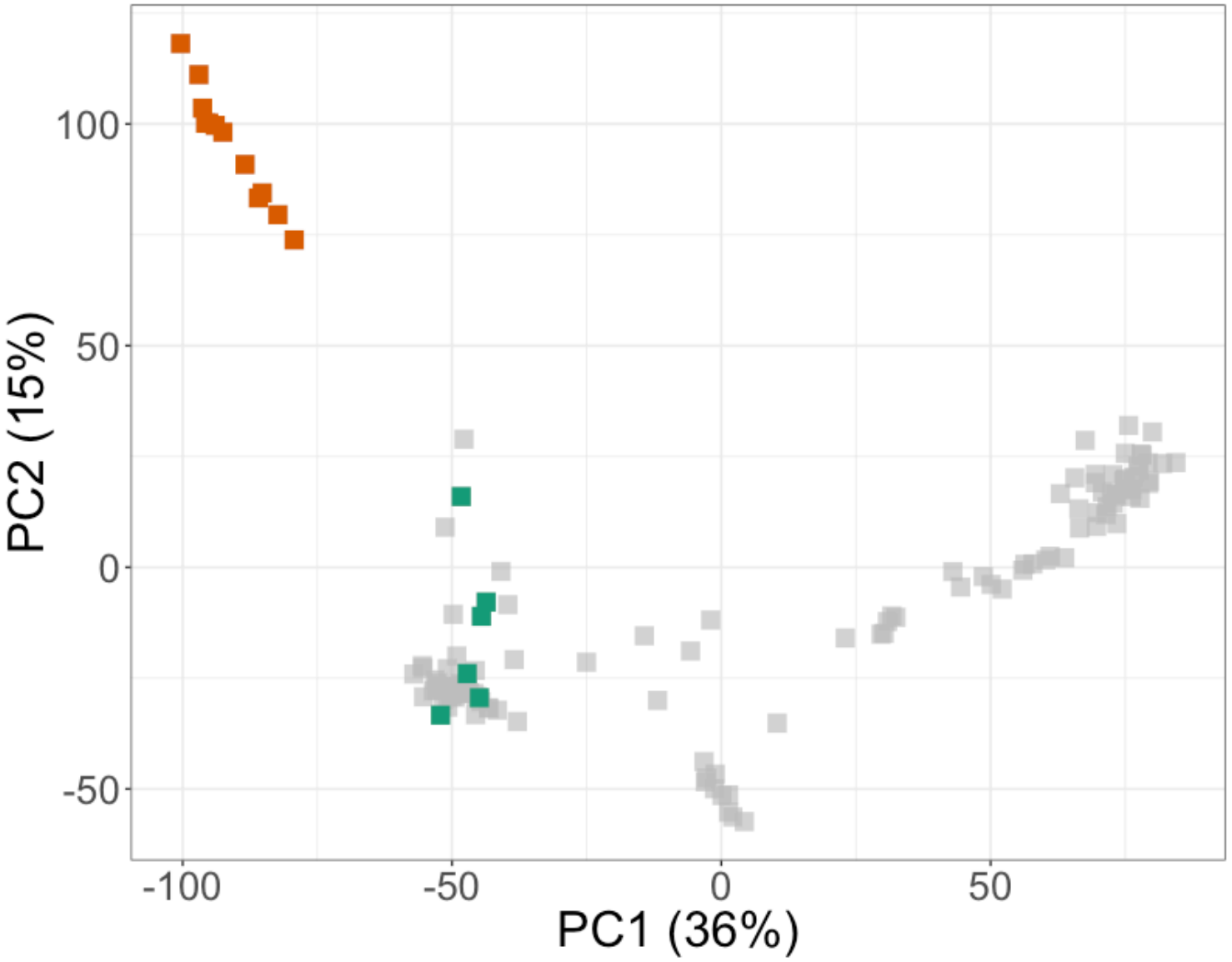



Group
- BRNCDT
- BRNNCC
- BRNPTM

*CG-DMRs*

| | n | Size |
|---|---|---|
| 5-group | 181,146 | 196.9 Mb |
| Basal ganglia | 16,866 | 24.0 Mb |

# Hippocampus: What the hell is going on?



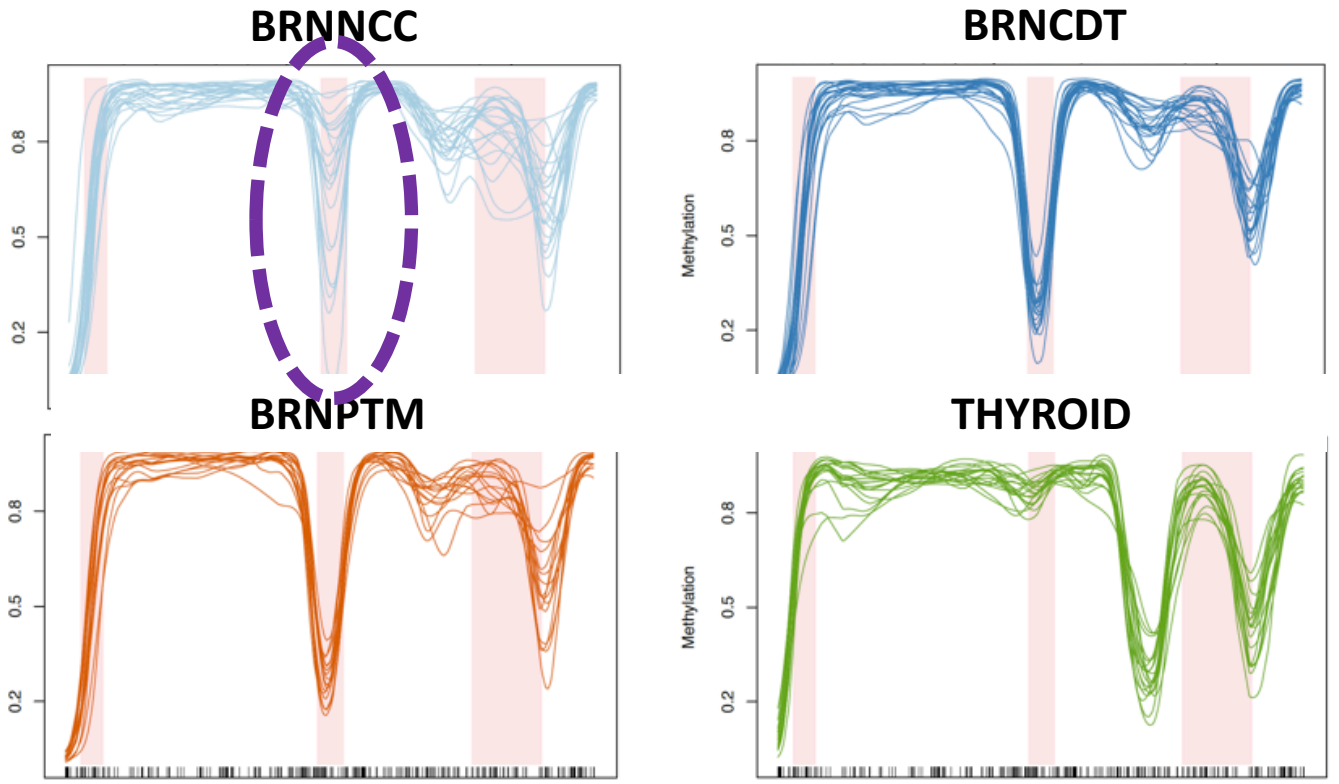

Group
- BRNHPP1
- BRNHPP2

*CG-DMRs*

| | n | Size |
|---|---|---|
| 5-group | 181,146 | 196.9 Mb |
| Basal ganglia | 16,866 | 24.0 Mb |
| Hippocampus | 11,702 | 24.4 Mb |

# Ongoing eGTEx analyses

- Complete analyses of CG-DMRs
- Identify CH-DMRs and analyse
- Stratified linkage disequilibrium score regression
  - Do BrainEpigenome results replicate?
  - What can brain region-specific DMRs tell us?
- **Variably methylated regions (VMRs)**
- Allele-specific methylation using phased GTEx genomes
- Use sorted data to deconvolute bulk brain samples
- Integration with other GTEx and eGTEx data



# Summary

- BrainEpigenome
  - FANS + WGBS reveals many brain region-specific CG-DMRs and CH-DMRs for NeuN+ (but not NeuN-) samples.
  - Neuronal CG-DMRs are enriched for heritability of several neurological, psychiatric, behavioral-cognitive phenotypes.
- eGTEx
  - More tissues + more replicates = huge increase in DMRs.
- The scale of these projects necessitated extensive improvements to computational methods and software engineering.
- There will still be **heaps of analyses on the table** after publication of initial eGTEx publication(s).
  - Get involved!

# Acknowledgements

Dr. Lindsay Rizzardi

Assoc. Prof. Kasper Hansen

Prof. Andy Feinberg

**Sequencing gurus**: Rakel Tryggvadóttir, Adrian Idrizi, Colin Callahan
**ATAC-seq experiments**: Varenka Rodriguez DiBlasi
**Flow sorting**: Hao Zhang and Hopkins Flow Facility
**Funding**: eGTEx (U01MH104393), CFAR (5P30AI094189-04, 1S10OD016315-01, and 1S10RR13777001), *AGTA Travel Award*
**Donors and families:** NIH NeuroBioBank at the University of Maryland & University of Pittsburgh

# Links

## Papers

*Rizzardi, L*. Hickey, P.F.*, et al.* **Neuronal brain region-specific DNA methylation and chromatin accessibility are associated with neuropsychiatric disease heritability**. *bioRxiv* (2017), doi:10.1101/120386 (in press, Nature Neuroscience)

*eGTEx Project* **Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease**. *Nature Genetics (2017),* doi: 10.1038/ng.3969

## Genome Browser

www.bit.ly/BrainEpigenomeHub

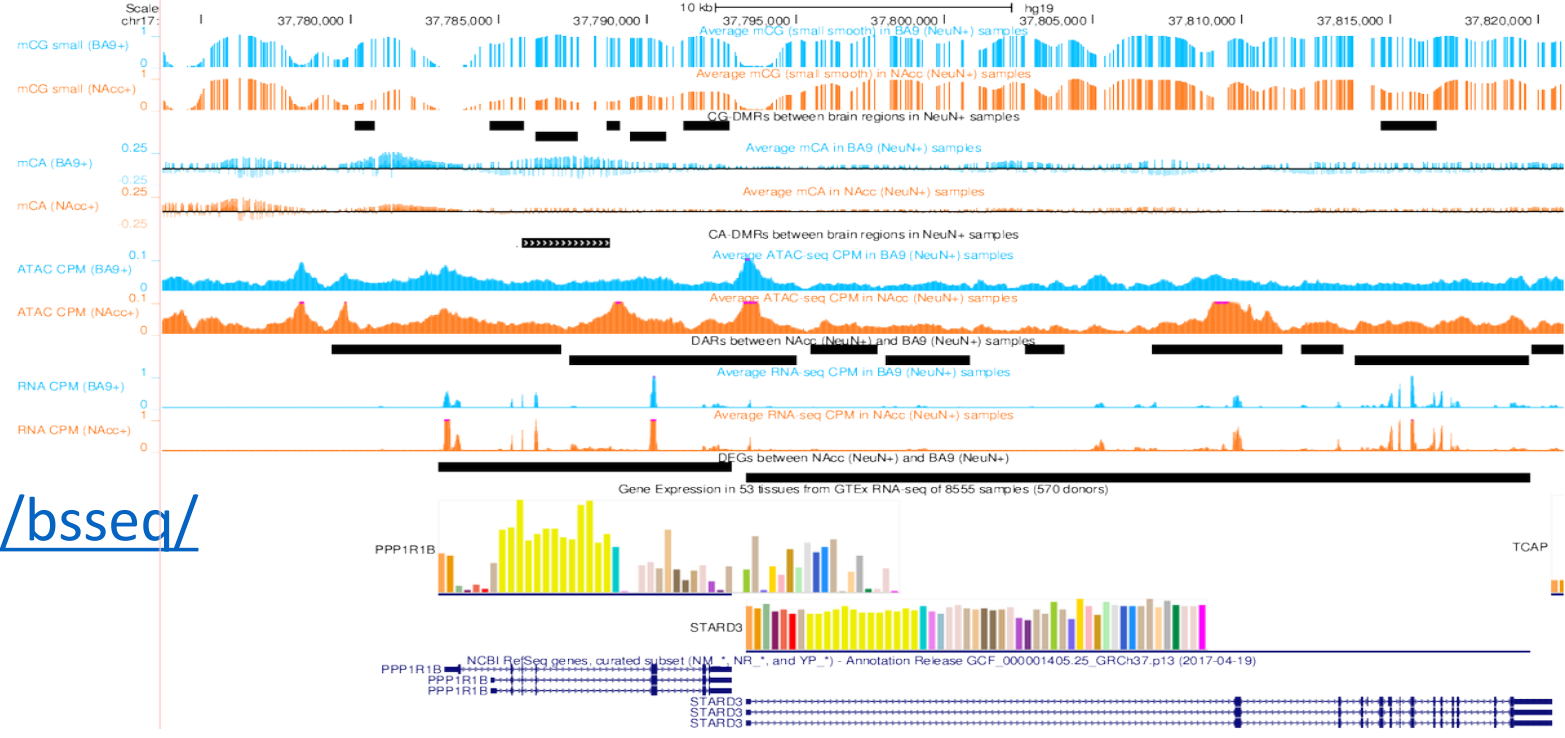
## Slides

www.bit.ly/AGTA2018

## Software

http://bioconductor.org/packages/bsseq/

@PeteHaitch



# Bonus slides

# eGTEx capture bisulfite-sequencing study

- Aim: Study genetic influence on DNA methylation in human brain
- Assay: Targeting 46 Mb (1 million CpGs) with Roche NimbleGen capture
  - 55% of CpGs not captured by microarrays or other targeted panels
  - CG-DMRs
    - Neuronal (BrainEpigenome and eGTEx)
    - NeuN+ vs. NeuN- (BrainEpigenome)
    - GABAergic vs. glutamatergic[1]
  - CG-VMRs (eGTEx)
  - Haplotype-dependent allele-specific DMRs and meQTLs[2]
  - Fetal brain meQTLs[3]
  - 'Epigenetic age' CpGs[4]
- Samples: > 100 donors (BRNCTXB, BRNCDT, BRNNCC, BRNHPP, and THYROID)

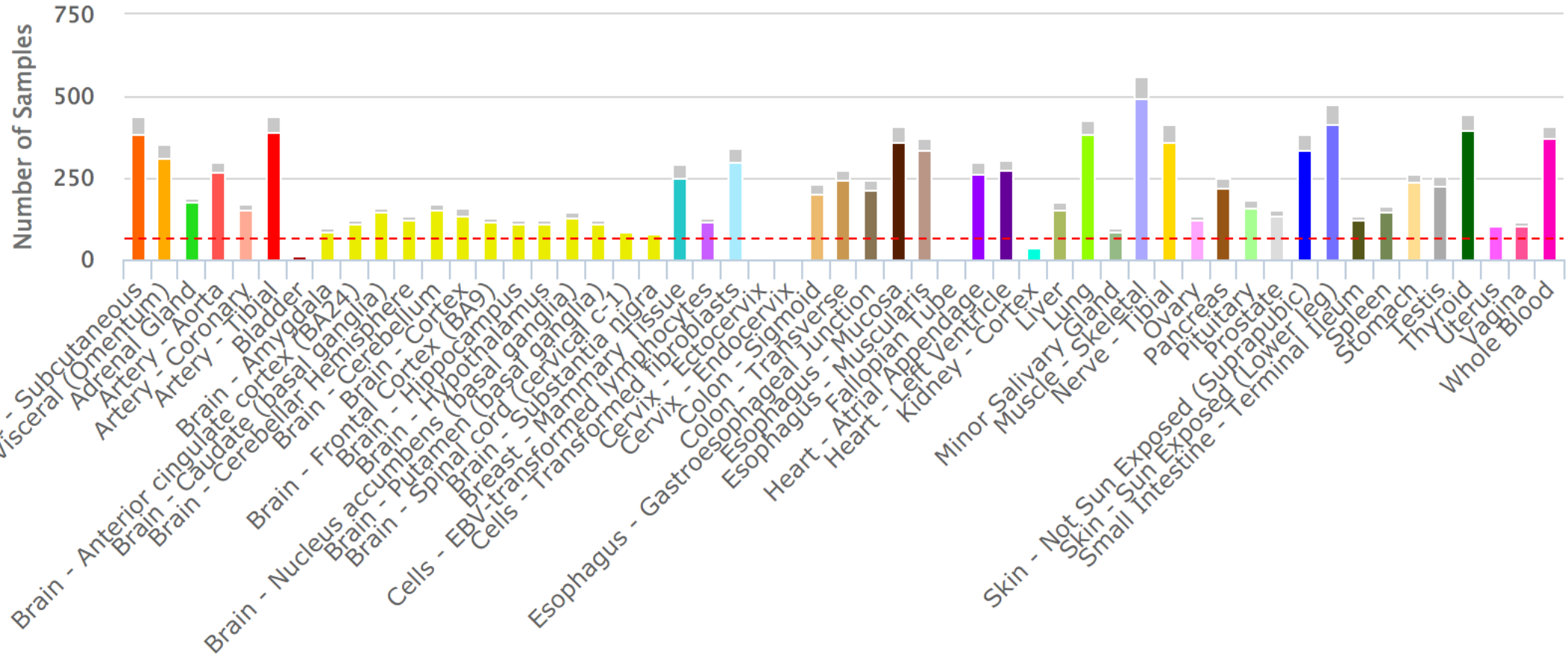
[1]Dracheva et al., *unpublished*

[2]Do, C. *et al.* **Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA Methylation**. *Am. J. Hum. Genet.* (2016)

[3]Court, F. *et al.* **Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment**. *Genome Res.* (2014)

[4]Horvath, S. **DNA methylation age of human tissues and cell types.** *Genome Biol.* (2013)

# GTEx -> eGTEx




[1]GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348,** 648–660 (2015).

[2]https://gtexportal.org/home/tissueSummaryPage

[3]eGTEx Project. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nat. Genet.* **49,** 1664–1670 (2017).