Accelerating Medicines Partnership - Alzheimer's Disease

[www.synapse.org/ampad]

Thanneer M Perumal

Senior Research Scientist

thanneer.perumal@sagebase.org

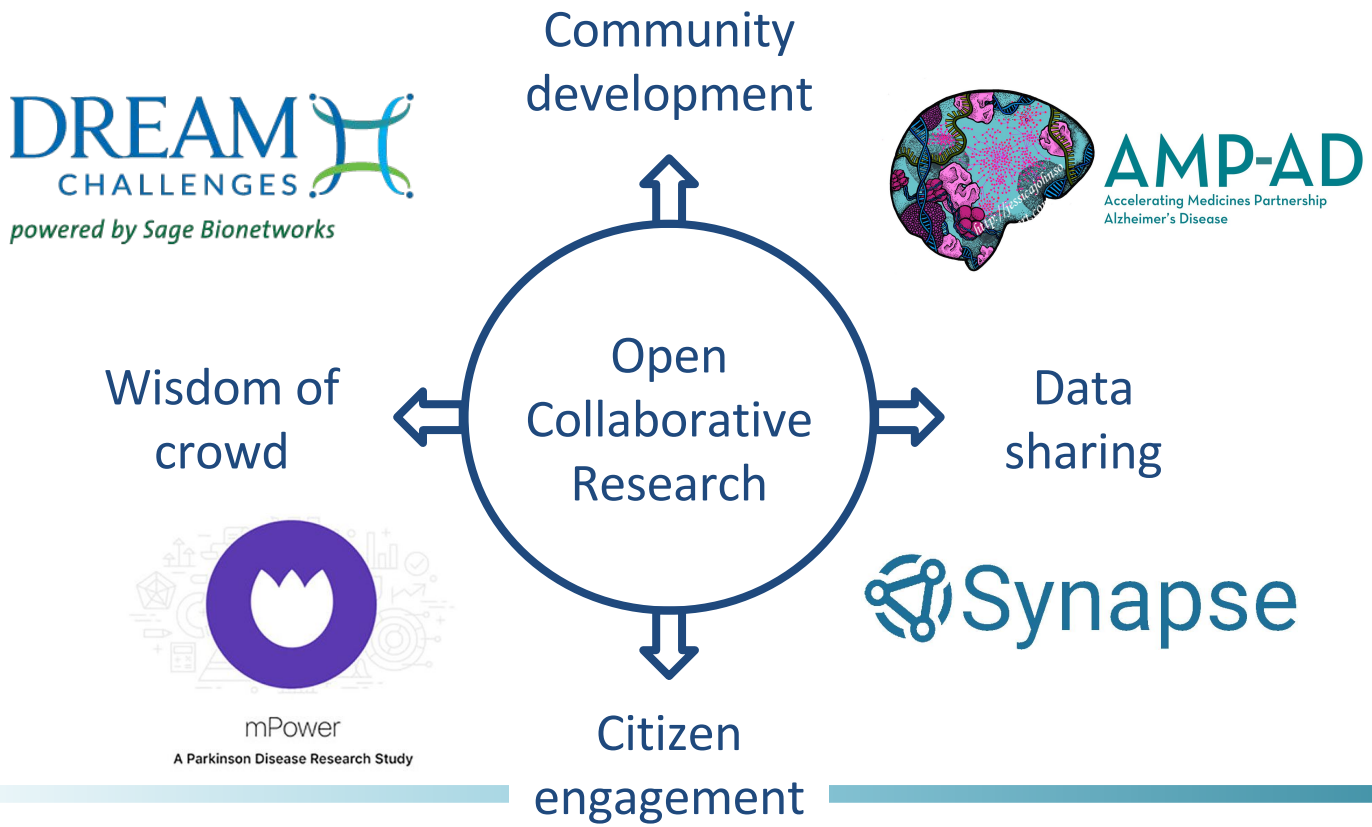
AMP-AD

Accelerating Medicines Partnership
Alzheimer's Disease

Sage BIONETWORKS

Sage Bionetworks

We are a Seattle based non-profit research organization promoting open systems, incentives, and norms to redefine how complex biological data is gathered, shared, and used



Accelerating Medicines Partnership

The New York Times

HEALTH

An Unusual Partnership to Tackle Stubborn Diseases

By GINA KOLATA FEB. 4, 2014

A public/private partnership between NIH, 10 biopharmaceutical companies, and several non-profit organisations

AMP-AD
Accelerating Medicines Partnership
Alzheimer's Disease

http://www.nytimes.com/2014/02/05/health/nih-joins-drug-makers-and-nonprofits-on-stubborn-diseases.html?_r=0

Accelerating Medicines Partnership

($ Millions)	Total Project	Total NIH	Total Industry
Alzheimer's Disease	92.5	69.6	22.9
Type 2 Diabetes	58.4	30.4	28
Rheumatoid Arthritis	41.6	20.9	20.7
Total	**192.5**	**120.9**	**71.6**

https://www.nia.nih.gov/alzheimers/amp-ad

Accelerating Medicines Partnership

BWH, RUSH	Genomics, Transcriptomics	iPS neurons and astrocytes, iPS compound screens
Icahn School of Medicine at Mount Sinai	Genomics, Transcriptomics	iPS neurons and astrocytes, Mouse hippocampal slices, Drosophila model
UFHealth University of Florida Health, Mayo Clinic, Institute for Systems Biology	Genomics, Transcriptomics	aB and Tau mouse models, rAAV delivered models
Emory University	Proteomics	Cell culture models, Mouse models, Drosophila model
Duke University	Metabolomics	ApoE4 TR and APP/PSI/ApoE4 mouse models
Harvard Medical School	Genomics, Transcriptomics	Studying the healthy human brain

Accelerating Medicines Partnership

Target Discovery | Target Discovery | Target Discovery | Target Discovery

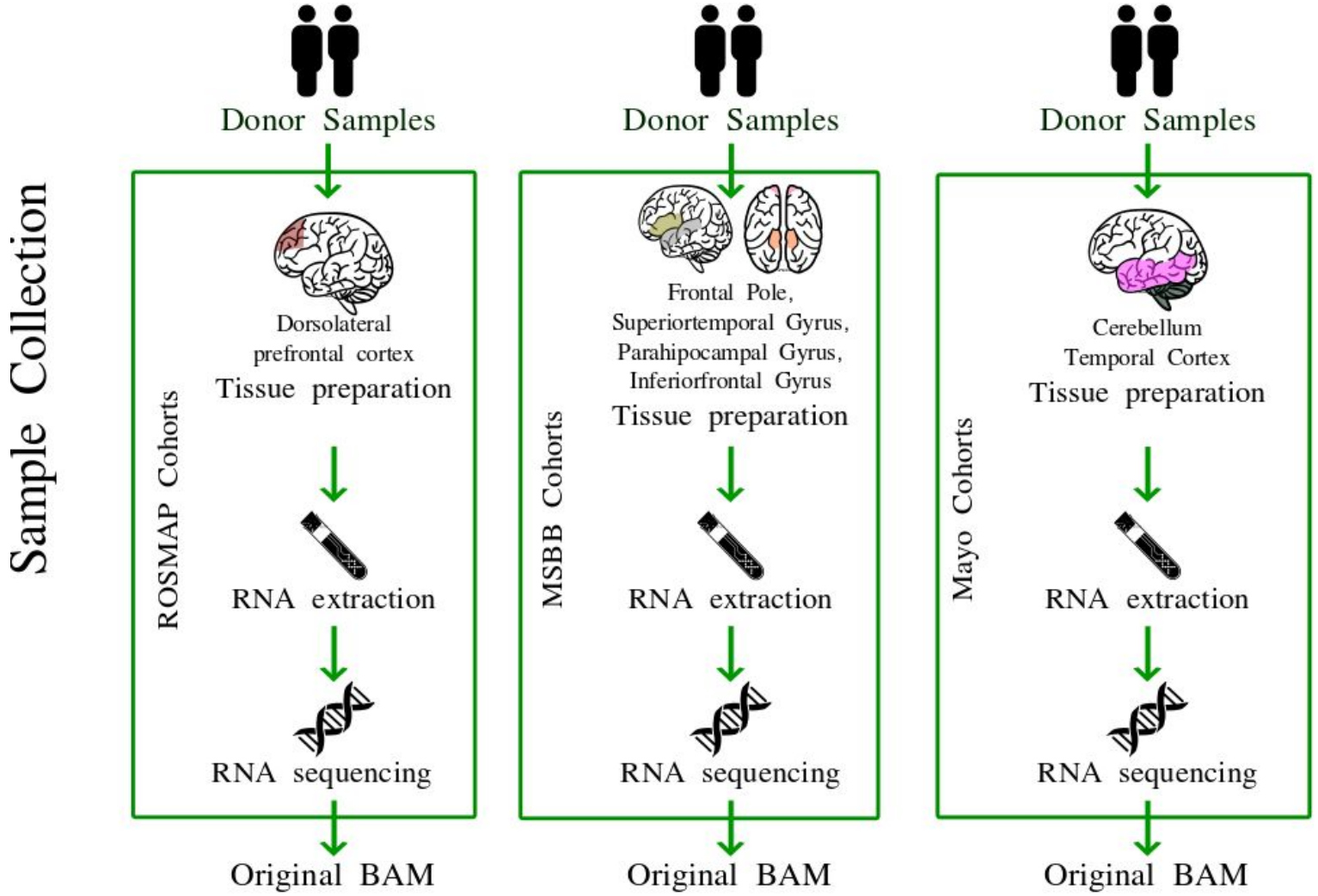
Sage BIONETWORKS

Coordinated sharing of early-phase target identification insights

Target Validation | Target Validation | Target Validation | Target Validation

Quarterly public release of data and analysis results through Synapse: www.synapse.org/ampad

AMP-AD RNASeq Sample Overview



of samples (based on harmonised case-control)

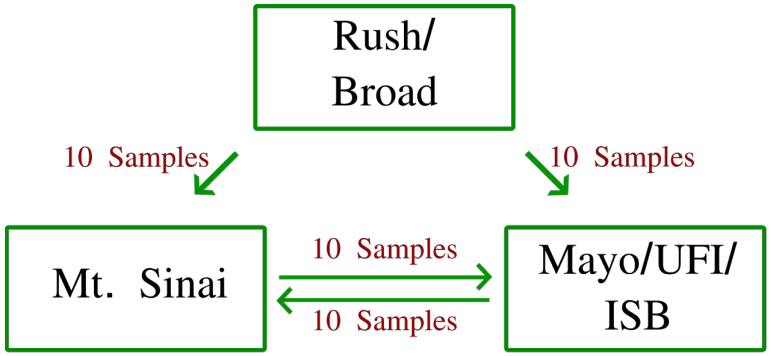
Source	Region	Definition		# of Samples	
		CONTROL	AD	CONTROL	AD
ROSMAP	DLPFC	cogdx=1, Braak <=3, CERAD >=3 and no other pathology	cogdx=4, Braak >=4, CERAD <=2	86	155
MSBB	FP	CDR<=0.5, Braak <=3, NP.1 <= 1	CDR>=1, Braak >=4, NP.1 >= 2	45	90
	IFG			37	79
	PHG			40	65
	STG			37	86
MAYO	CER	No AD, Braak <= 3, CERAD <=1 and no other pathology	Definite AD from NINCDS-ADRDA criteria, Braak >= 4	74	79
	TCX			73	80
Total				**392**	**634**

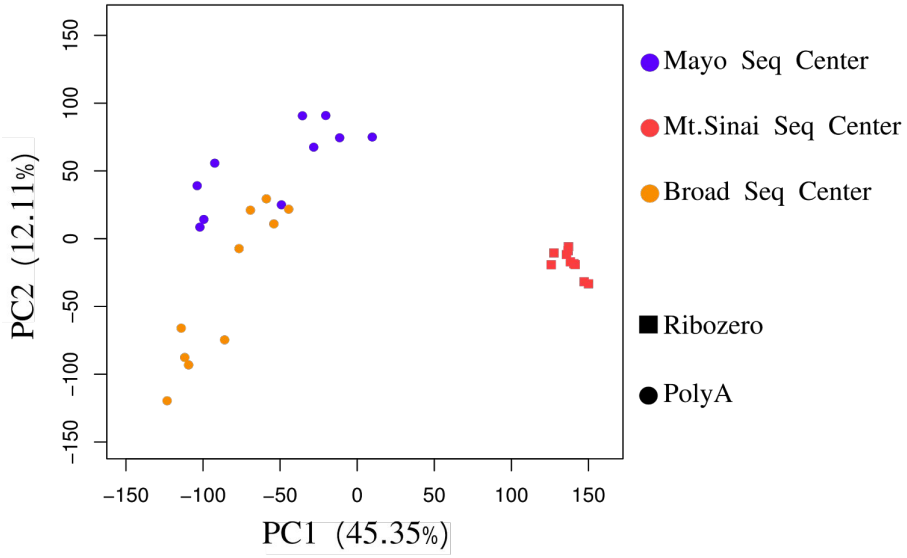
AMP-AD RNASeq Sample Overview

Source	# of Individuals	Brain Regions	Library Preparation	Read Length	Depth of Sequencing
ROSMAP	634	 Dorsolateral prefrontal cortex	Poly-A Selection	101 bp Paired End	50-100M per library
MSBB	301	 Frontal Pole, Superiortemporal Gyrus, Parahipocampal Gyrus, Inferiorfrontal Gyrus	RiboZero	100 bp Single End	120M per library
MAYO	303	 Cerebellum Temporal Cortex	Poly-A Selection	100 bp Paired End	120M per library

RNASeq Sample SWAP

- ***Objective:***
  - To quantify technical variations across different sequencing centers
- ***Design:***
  - 10 Samples from each cohort was re-sequenced across all three centers





- Top 2 principal components of gene expression from 10 ROSMAP samples that were sequenced at all three centers: Broad, Mt.Sinai and Mayo
  - PC1: Differences between Ribozero vs PolyA preparation
  - PC2: Sequencing centers

Sample SWAP analysis performed by Dr. Kristen Dang: https://www.synapse.org/#!Synapse:syn4584500/wiki/390727

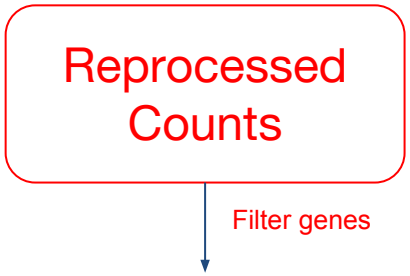
RNASeq Reprocessing*

- Goal:
  - To create a ***uniformly processed*** RNAseq dataset across the three largest AMP-AD contributed studies (ROSMAP/MSBB/MayoRNAseq)

- Method:
  - FASTQ generation
    - Conversion of BAMs to FASTQs using Picard.
  - Alignment
    - Re-alignment of FASTQ reads using STAR.
    - Computing alignment metrics using Picard.
  - Gene counting
    - Counting reads per gene using STAR

* reprocessing performed by Dr. Kristen Dang & Dr. James Eddy: https://www.synapse.org/#!Synapse:syn9702085

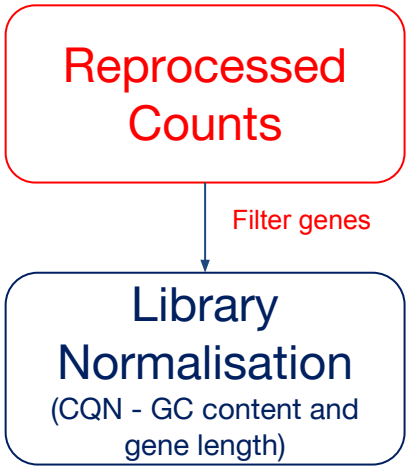
Gene Filtering



- Gene Filtering
  - Filter genes that have less than 1 CPM in more than 50% of samples for each brain region times diagnosis category

	ROSMAP	MSBB	MAYO
# of Genes	15582	16348	17003
Fraction of antisense	0.03	0.04	0.05
Fraction of lincRNA	0.03	0.04	0.05
Fraction of protein coding	0.87	0.85	0.84
Fraction of transcribed unprocessed pseudogene	0.01	0.01	0.01

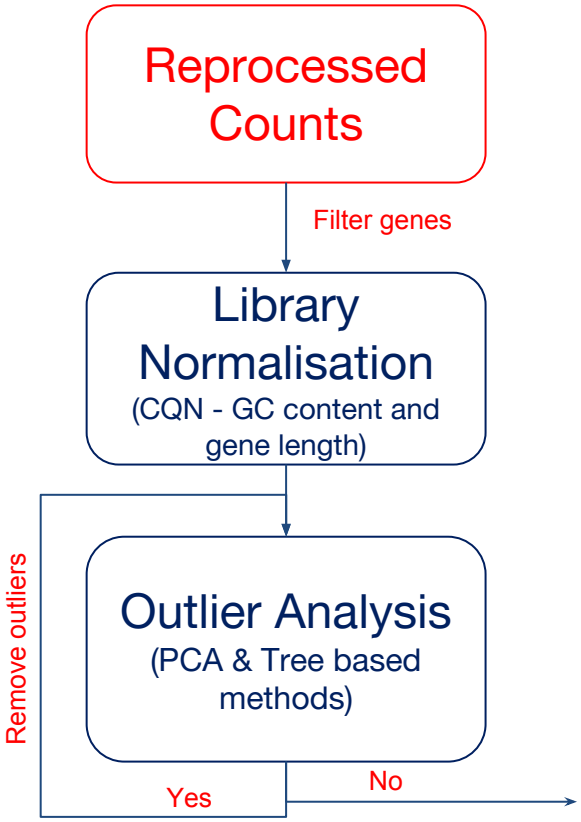
Library Normalisation



- Gene Filtering
  - Filter genes that have less than 1 CPM in more than 50% of samples for each brain region times diagnosis category
- Library Normalisation (using CQN*)
  - Conditional quantile normalisation accounting for gene length and GC content

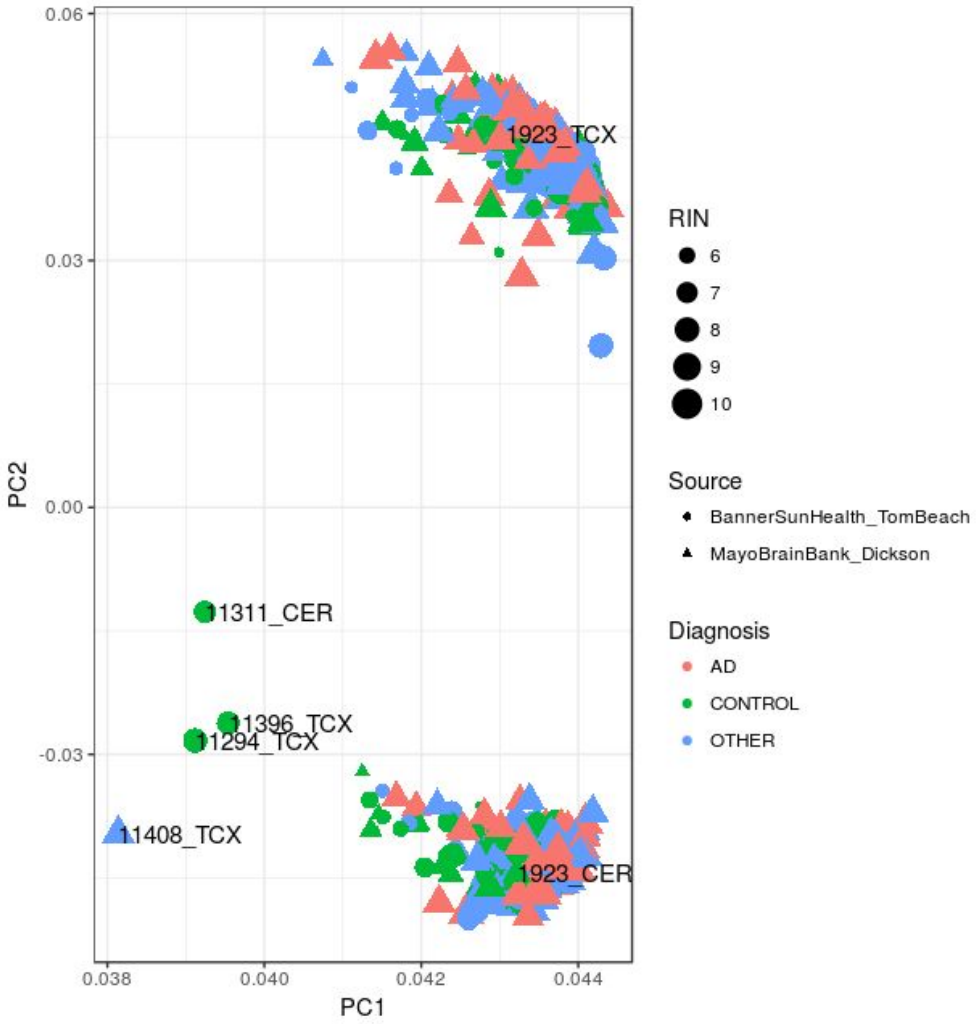
*Hansen KD, Irizarry RA and Wu Z (2012). "Removing technical variability in RNA-seq data using conditional quantile normalization." *Biostatistics*, **13**(2), pp. 204–216.

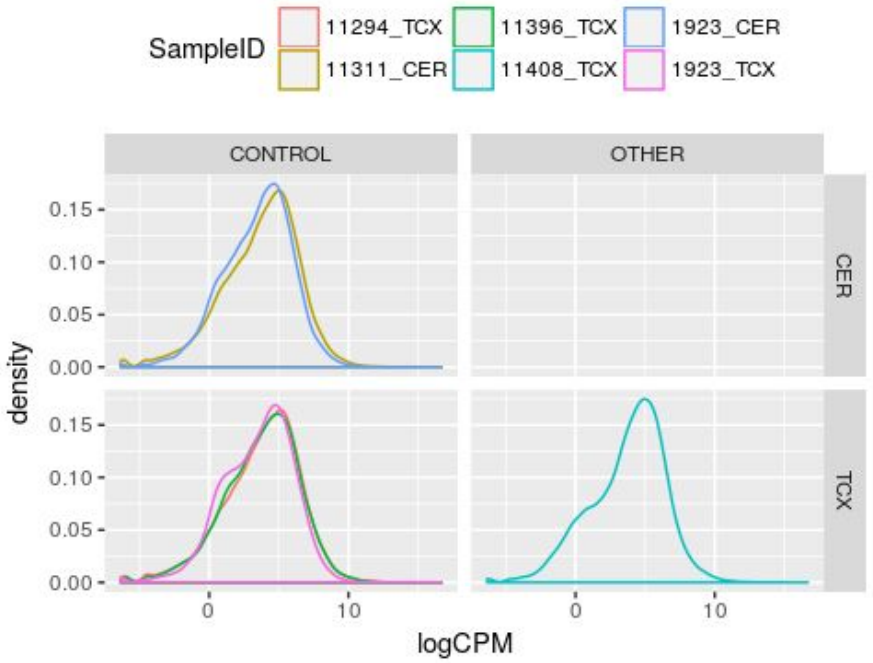
Iterative Outlier Analysis



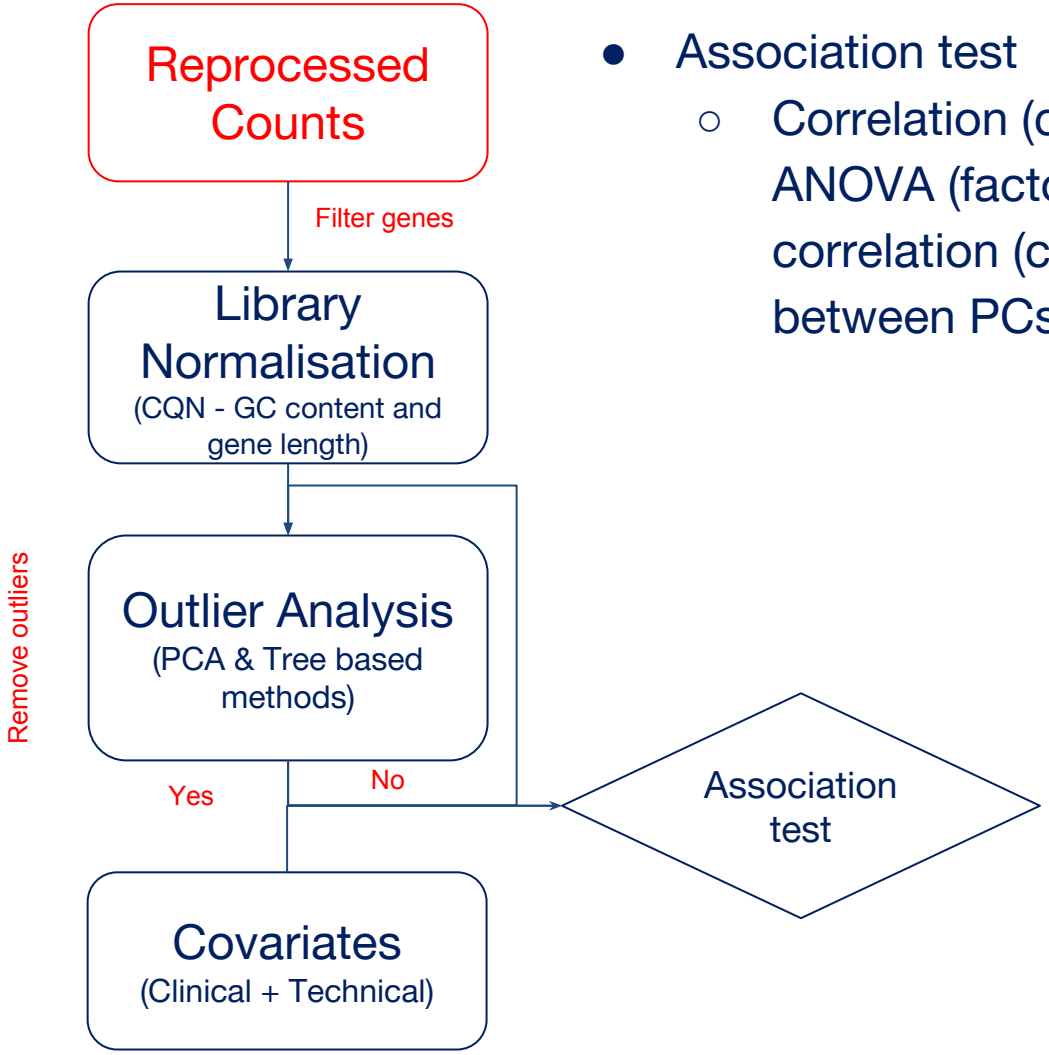
- Gene Filtering
  - Filter genes that have less than 1 CPM in more than 50% of samples for each brain region times diagnosis category
- Library Normalisation (using CQN)
  - Conditional quantile normalisation accounting for gene length and GC content
- Iterative outlier detection
  - PCA and euclidean distance based dendrogram trees were used to identify outliers

Iterative Outlier Analysis: MAYO Samples



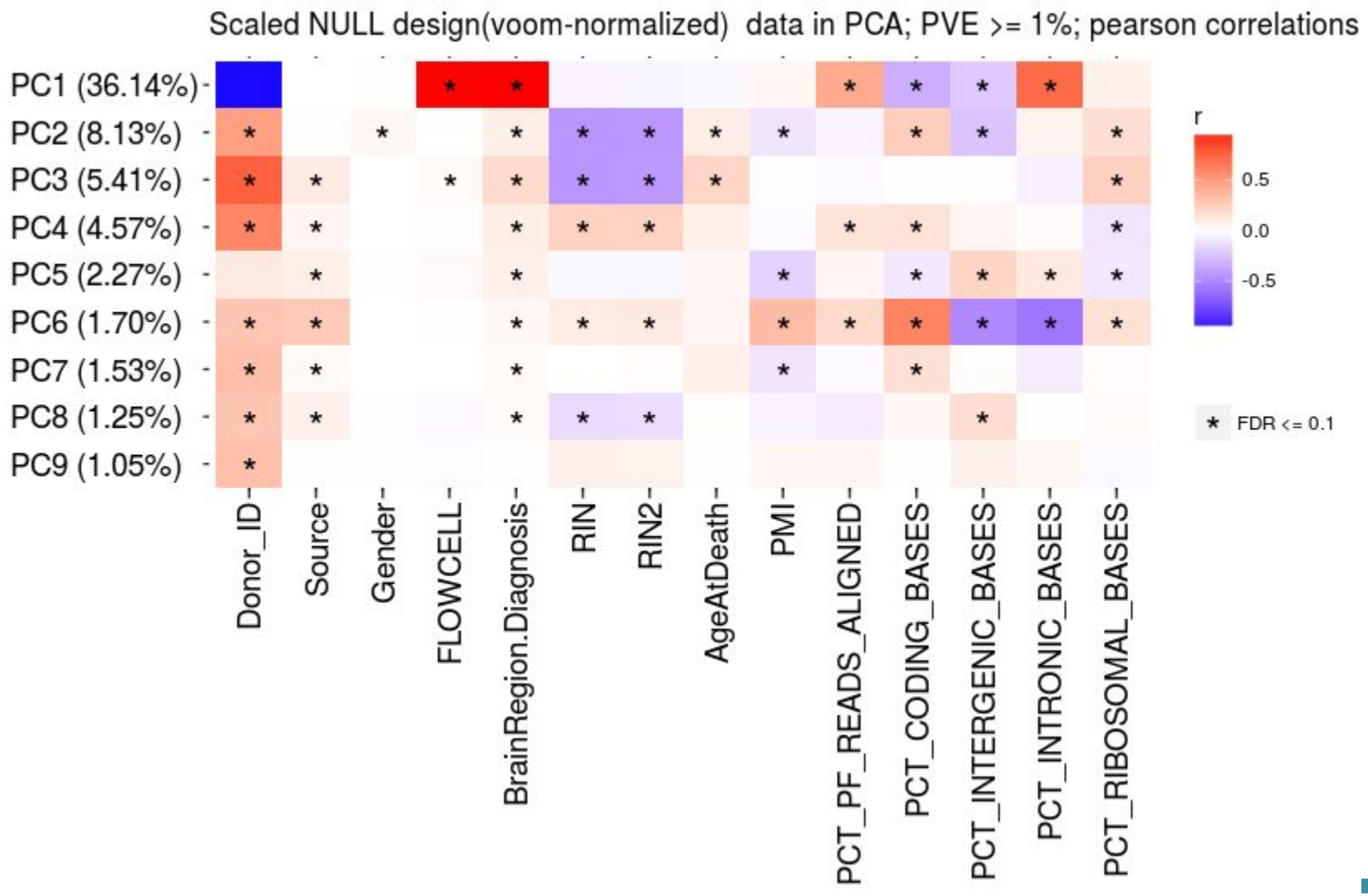


Association b/w Expression and Covariates

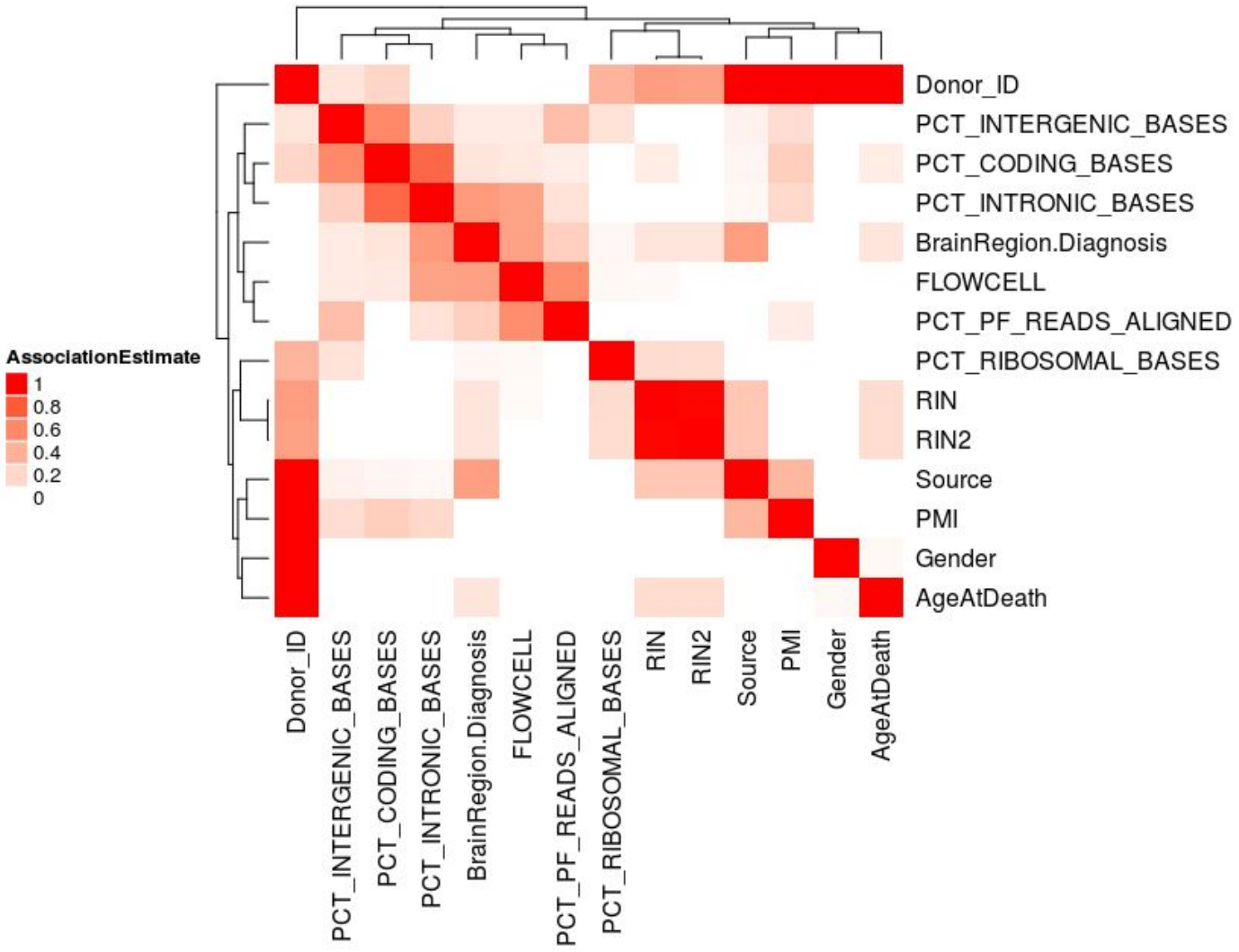


- Association test
  - Correlation (continuous - continuous), one-way ANOVA (factor-factor) and/or Inter class correlation (continuous - factor) was calculated between PCs of expression and covariates

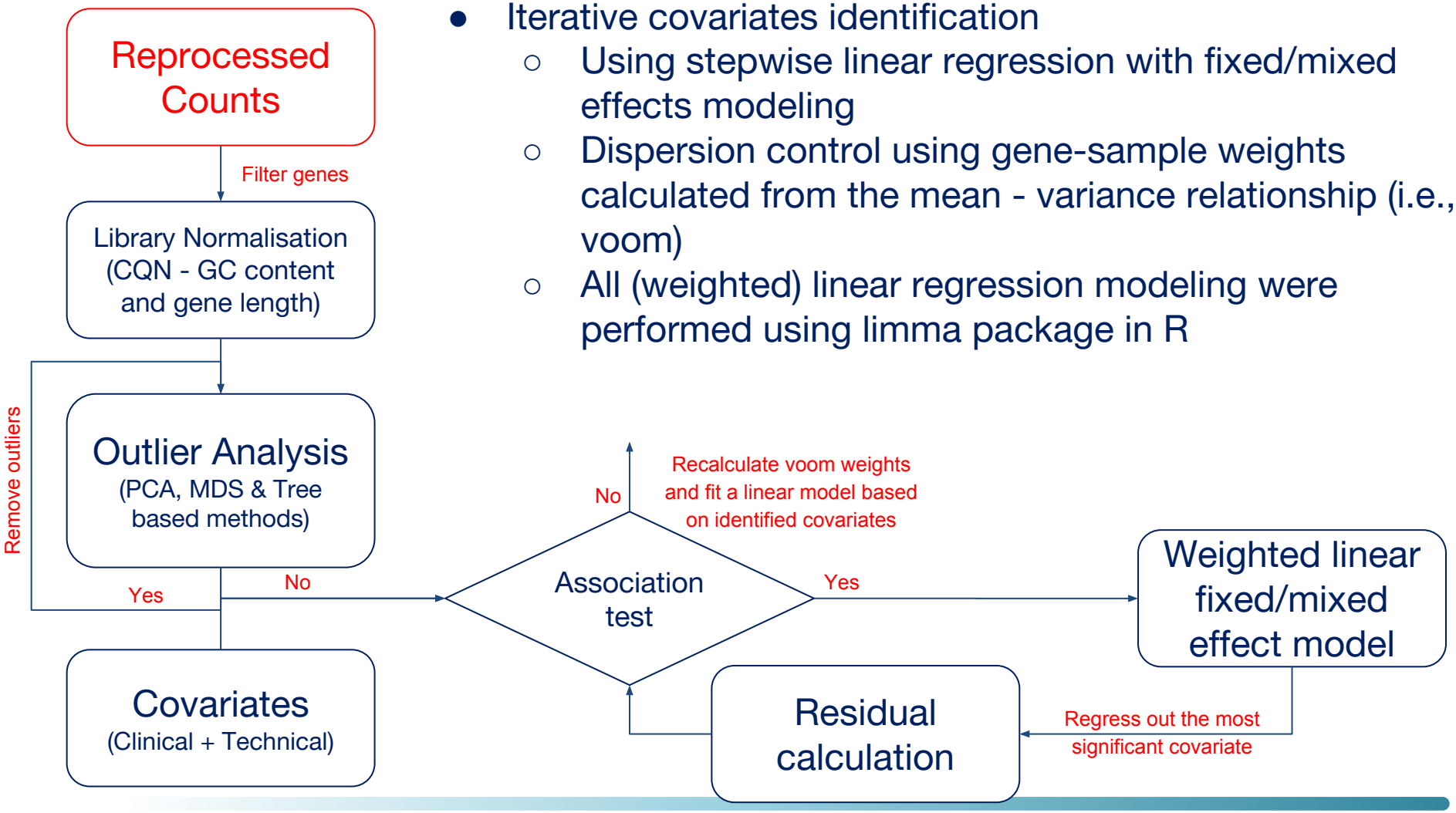
Expression Association with Covariates in MAYO Cohort



Association b/w Covariates

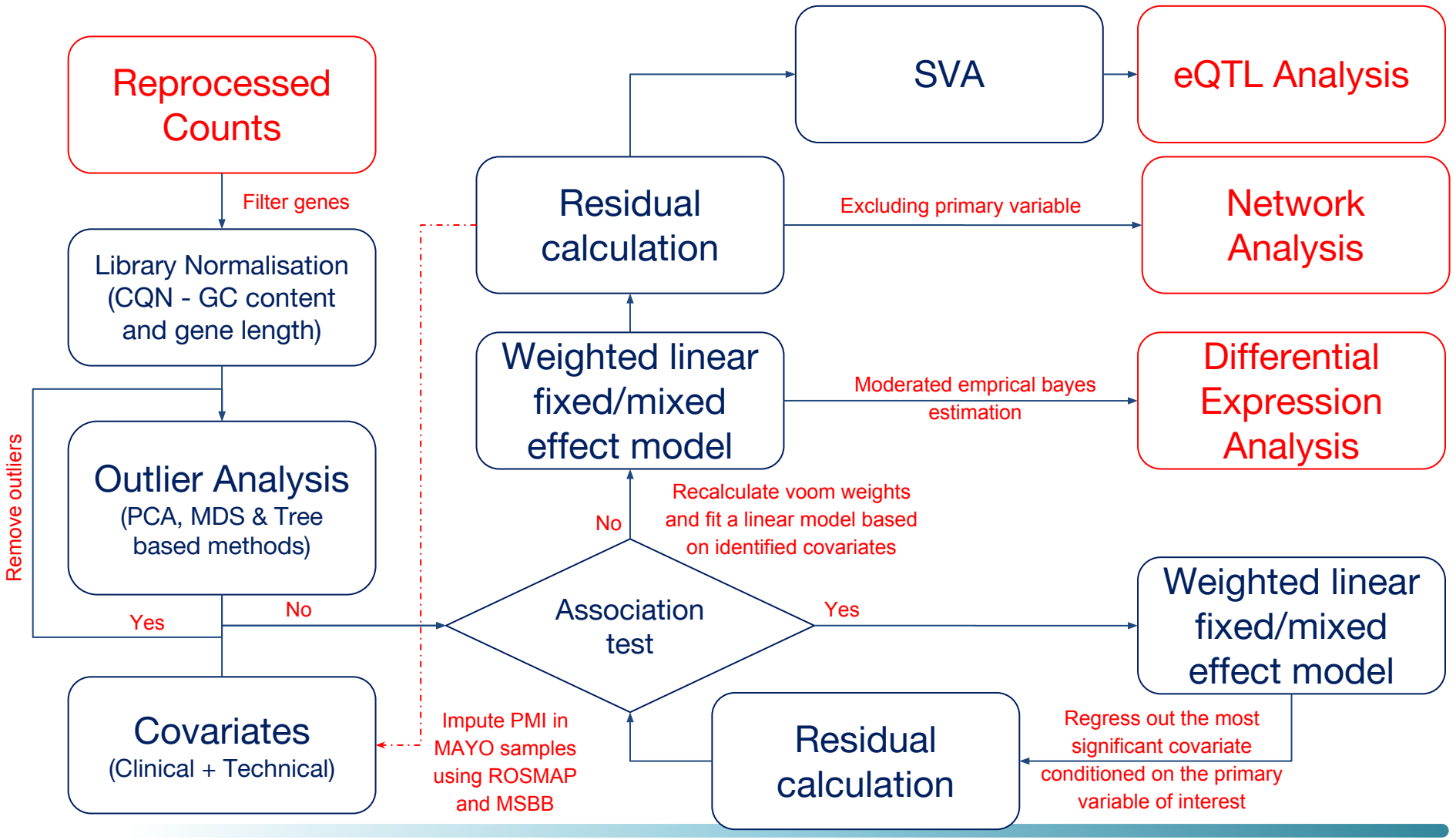


Iterative Covariates Identification

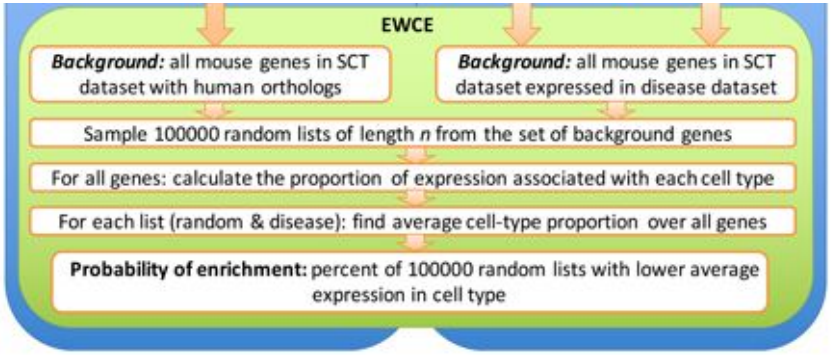


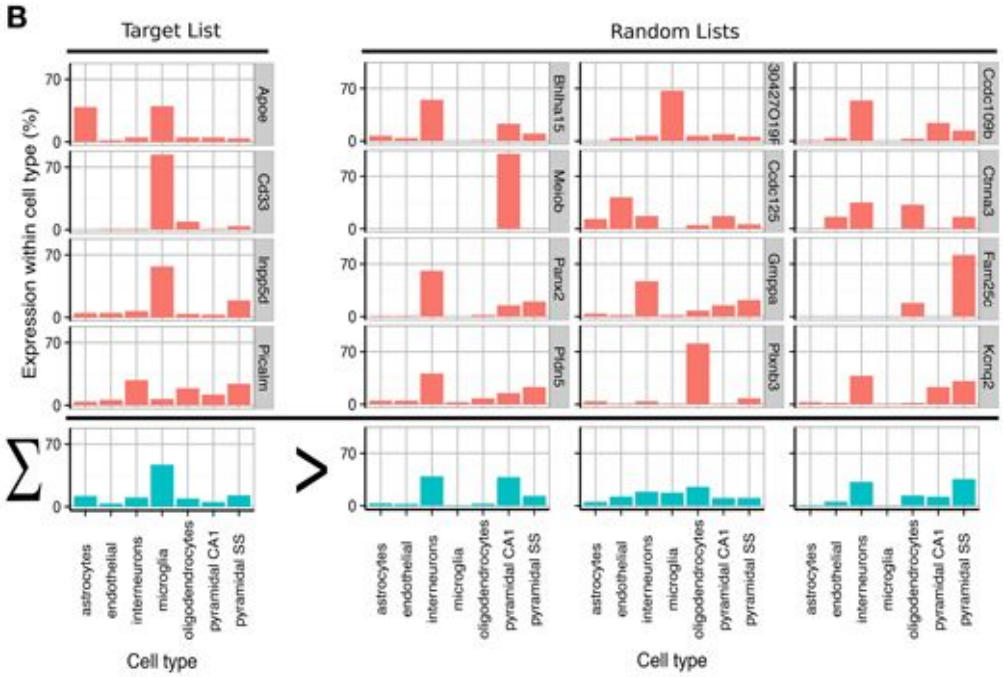
- Iterative covariates identification
  - Using stepwise linear regression with fixed/mixed effects modeling
  - Dispersion control using gene-sample weights calculated from the mean - variance relationship (i.e., voom)
  - All (weighted) linear regression modeling were performed using limma package in R

Covariate Adjusted Data



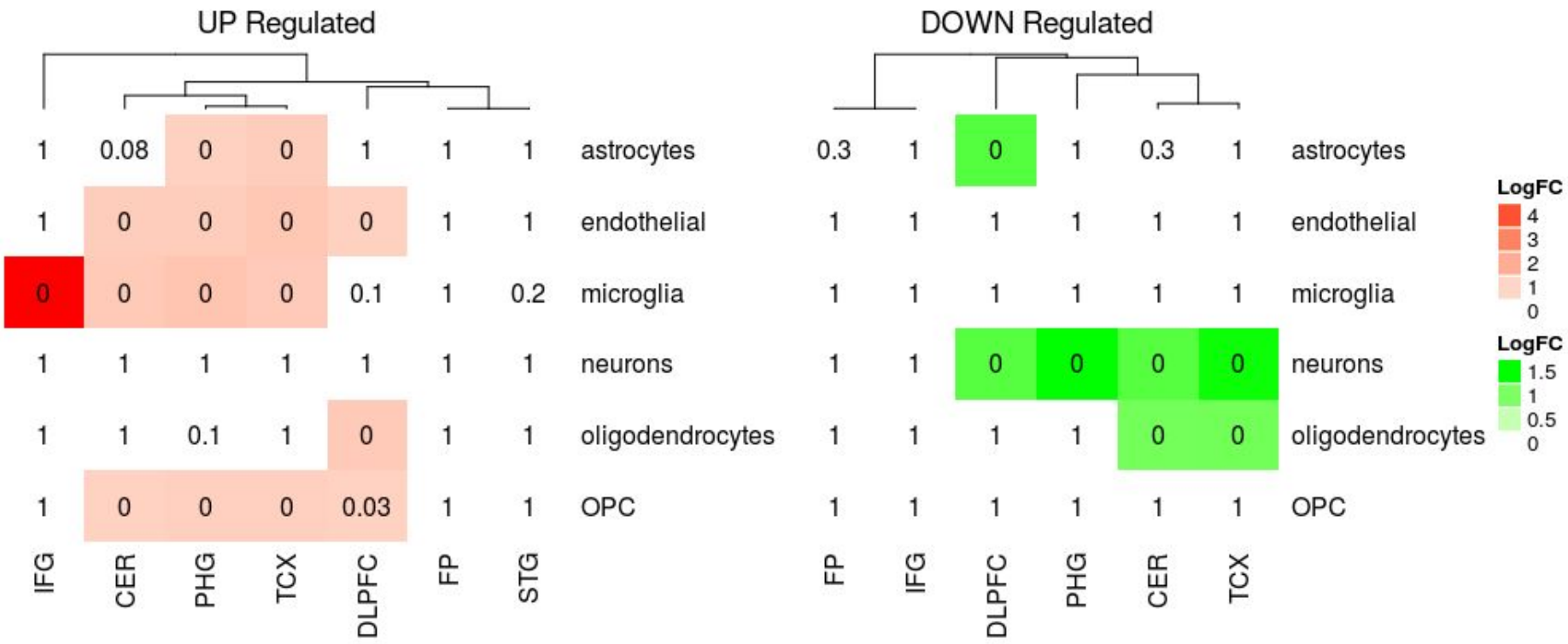
Expression Weighted Cell Type Enrichment*





*N.G. SKENE, S.G.N. GRANT (2016) Expression weighted cell type enrichments reveal genetic and cellular nature of major brain disorders. Frontiers in Neuroscience.

Expression Weighted Cell Type Enrichment*



*N.G. SKENE, S.G.N. GRANT (2016) Expression weighted cell type enrichments reveal genetic and cellular nature of major brain disorders. Frontiers in Neuroscience.

Upcoming Challenges

GA4GH/DREAM Workflow Execution Challenge

Launches July 5 (Pre-registration is open)

The goal of this challenge is to evaluate systems and platforms for executing portable analysis workflows in the interest of developing common standards and best practices. Participants will run high quality genomics workflows in their system of choice to assess portability and reproducibility in a concrete way.

Parkinson's Disease Digital Biomarker DREAM Challenge

Launching July 6, 2017 (Pre-registration is open)

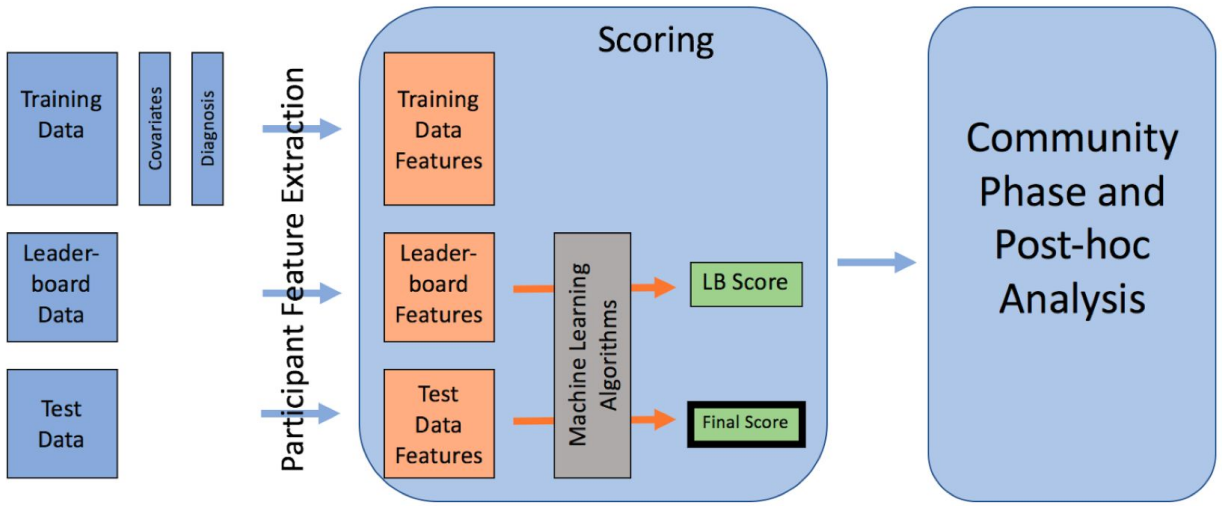
Using data collected through two Parkinson's Disease mobile research studies, the goal of this challenge is to identify the best methods for processing mobile sensor data in order to distinguish gait and motor differences between Parkinson's Disease patients and controls.

http://dreamchallenges.org/upcoming-challenges/

Parkinson's Disease Digital Biomarker DREAM Challenge

ENABLED BY

HARVARD UNIVERSITY · THE MICHAEL J. FOX FOUNDATION FOR PARKINSON'S RESEARCH · Robert Wood Johnson Foundation · Sage · verily · University of Rochester · EARLY SIGNAL



- ✓ Launched: 6 July 2017
- ✓ Close: October 2017

https://www.synapse.org/#!Synapse:syn8717496/wiki/422884

Acknowledgements

Sage Bionetworks Team

Kristen Dang

James Eddy

Ben Logsdon

Solly Sieberts

Mette Peters

Lara Mangravite

Larsson Omberg

Mt. Sinai Team

Minghui Wang

Bin Zhang

Eric Schadt

Rush Team

Jishu Xu

Philip De Jager

David Bennett

UFL-Mayo-ISB Team

Xue Wang

Mariet Allen

Nilufer Taner

Todd Golde

Nathan Price

Lilly Team

Philip J Ebert

John Calley

Yuhao Lin

Corey James

David Collier

Funding

We would like to thank the NIH for their financial support: U01 AG046152, U01 AG046139, U01 AG046161, U01 AG046170, BDDS (1U54EB020406-01)

AMP-AD Differential Expression Analysis Working Group Members

AMP-AD Network Analysis Working Group Members

AMP-AD RNASeq Reprocessing Working Group Members

And many others...

Thank You

thanneer.perumal@sagebase.org

AMP-AD
Accelerating Medicines Partnership
Alzheimer's Disease

Sage
BIONETWORKS