Assessing the causal association of mtDNAcn with Alzheimer's disease

Dr. Shea Andrews

28 May, 2020

Contents

Abstract

Increasing evidence has implicated mitochondrial dysfunction in Alzheimer's Disease (AD). As AD features altered mitochondrial function, this suggests that therapeutics strategies aimed at preventing declines in mitochondrial function may modify the disease course in AD. However, it is unclear whether mitochondrial dysfunction causes, mediates, or is a by-product of AD pathogenesis. As mitochondria contain their own DNA outside of the nuclear genome, with every cell having between 100-10,000 copies of mitochondrial DNA, mitochondrial DNA copy number (mtDNA-CN) can be used as a surrogate measure of mitochondrial function. The overall objective of this research program is to evaluate whether mitochondrial dysfunction plays a causal role in AD pathogenesis. Our central hypothesis is that lower mtDNA-CN – indicative of mitochondrial dysfunction – will be associated with increased risk of AD. This study will disentangle the causal role of mitochondrial dysfunction in AD using traditional epidemiological approaches, polygenic risk scoring (PRS) and Mendelian randomization (MR). PRS are a measure of an individual's genetic propensity to a trait and can be used to evaluate the genetic overlap between two traits by testing whether the PRS of one trait predicts another trait, while MR uses genetic variants to estimate the causal effect of risk factors on disease outcomes. In the first aim, we will calculate in mtDNA-CN in AD cases and controls and evaluate the association between mtDNA-CN and AD. In the second aim, we will construct a PRS for mtDNA-CN and determine if genetically predicted mtDNA-CN is associated with AD outcomes. In the final aim, we will use MR to evaluate the causal effect of mtDNA-CN on AD outcomes and the causal effect of AD on mtDNA-CN. By establishing if mitochondrial dysfunction has a causal role in AD pathogenesis, this study will provide evidence regarding the utility of mitochondrial therapeutic strategies in AD.

Chapter 1

Introduction

1.1 Neuropathological Confirmed AD

There is consensus to disentangle the clinicopathologic term “Alzheimer’s disease” from AD neuropathologic change. The former refers to clinical signs and symptoms of cognitive and behavioral changes that are typical for patients who have substantial AD neuropathologic change,and is the focus of recent NIA–AA-sponsored consensus re-ports on three defined stages in a clinical continuum that includes preclinical, mild cognitive impairment, anddementia. The latter refers to the presence and extentof neuropathologic changes of AD observed at autopsy, re-gardless of the clinical setting.

1.1.1 CERAD Criteria - 1991

Protocol provides neuropathologic definitions of such terms as “definite Alzheimer’s disease” (AD), “probable AD,” “possible AD,” and “normal brain” to indicate levels of diagnostic certainty (Mirra et al. (1991)). The CERAD Neuritic Plaque score forms the basis of later neuropathological difinitions.

Sections are tacken from:

- middle frontal gyrus
- superior and middle temporal gyri
- inferior parietal lobule
- hippocampus and entorhinal cortex
- midbrain

And scored as a semiquantitative measurment:

- Absent
- Sparese
- Moderate

- Frequent

An age-related plaque score is then determined by combining the age of the patient at death and the semiquantitative measure of plaques in the *most severely affected region of the neocortex*. This score is then intergrated with with clinical information the presence or absence of dementia.

1.1.2 NIA-Reagan Criteria - 1997

The modified NIA-Reagan diagnosis of Alzheimer's disease is based on consensus recommendations for postmortem diagnosis of Alzheimer's disease. The criteria rely on both neurofibrillary tangles (Braak) and neuritic plaques (CERAD). See NIA Working group consensus 1997 and corresponding editorial by Hyman et al 1997. Traditionaly, the criteria require a history of dementia, insofar as they were designed to help address the question of whether AD was the underlying cause of a patient's dementia.

- CERAD score is a semiquantitative measure of neuritic plaques
  - No neuritic plaques (C0)
  - Sparse/infrequent neuritic plaques (C1)
  - Moderate neuritic plaques (C2)
  - Frequent neuritic plaques (C3)
- Braak Stage is a semiquantitative measure of severity of neurofibrillary tangle (NFT) pathology.
  - no NFTs (B0)
  - stages I/II, with NFTs predominantly in en-torhinal cortex and closely related areas (B1)
  - stages III/IV, withNFTs more abundant in hippocampus and amygdala whileextending slightly into association cortex (B2)
  - stages V/VI,with NFTs widely distributed throughout the neocortex (B3)

CERAD / Braak	0	I/II	III/IV	V/VI
None	**Normal**	-	-	-
Sparse	-	**Low**	-	-
Moderate	-	-	**Intermediate**	-
Frequent	-	-	-	**High**

1.1.3 NIA-AA Criteria - 2012

The NIA-AA criteria updated and revised the 1997 NIA-Reagan criteria to recognize the pre-clinical stage of AD, enhance the assessment of AD to include amyloid accumulation as well as neurofibrillary change and neuritic plaques. Hyman et al 2012. The criteria relies on an 'ABC' score for AD neuropathologic change that incorporates histopathologic assessments of amyloid deposits (A - Thal phase), staging of neurofibrillary tangles (B - CERAD), and scoring of

neuritic plaques (C - Braak Stage). See Hyman et al 2012 for guidlines and Montine et al 2012 for a practial guide.

- Thal Phase is a semiquantitiative measure of the distribution of AB
  - phase 0 or no amyloid
  - phase 1 or isocortical
  - phase 2 or limbic
  - phase 3 or basal ganglia
  - phase 4 or basal forebrain and midbrain
  - phase 5 or pons/medulla oblongata and cerebellum

Thal	CERAD	Braak:	None or I/II (B0 or B1)	III/IV (B2)	V/VI (B3)
0 (A0)	None (C0)		Other§	Other§	Other§
1/2 (A1)	None - Sparse (C0 or C1)		Low	Low	Low¶
	Modearte - Frequent C2 or C3)		Low†	Intermediate	Intermediate¶
3 (A2)	Any C		Low†	Intermediate	Intermediate¶
4/5 (A3)	None - Sparse (C0 or C1)		Low†	Intermediate	Intermediate¶
	Modearte - Frequent C2 or C3)		Low†	Intermediate	High

§Medial temporal lobe NFTs in the absence of significant Ab or neuritic plaques occur in older people and may be seen in individuals without cognitiveimpairment, with mild impairment, or with cognitive impairment from causes other than AD. Consider other diseases when clinically or pathologically indicated.

¶Widespread NFTs with some Ab/amyloid plaques or limited neuritic plaques are relatively infrequent, and when they occur, other diseases, particularly-tauopathies, should be considered. Such cases may not fit easily into a specific Braak stage, which is intended for categorization of AD-type NFTs.

†Higher levels of Ab or neuritic plaques with low Braak stage should prompt consideration of contribution by comorbidities such as vascular brain injury,LBD, or HS. Also, consider additional sections as well as repeat or additional protocols to demonstrate other non-AD lesions

For individuals **without cognitive impairmentat** the time tissue was obtained, it is possible that AD neuropathologic change may predate onset ofsymptoms by years. For individuals **with cognitive impairmentat** the time tissue was obtained, “Intermediate” or “High” level (Table 2) of AD neuropathologic change should be considered adequate explanation of cognitive impairment or dementia. When “Low” level of AD neuropathologic change is observed in the setting of cognitive impairment, it is likely that other diseases are present. In all cases with cognitive impairment, regardless of the extent of AD neuropathologicchange, it is essential to determine the presence or absence, as well as extent, of other disease(s) that might have contributed to the clinical deficits.

Possibility that Thal amyloid stages do not substantially contribute to predicting antemortem cognition compared to CERAD neuritic plaque scores and

Braak NFT stages Serrano-Pozo et al 2016.

Chapter 2

Methods

This section describes the general methods used for calling mitochondrial haplogroups, estimating mtDNAcn and the cohorts used in the analysis.

2.1 Haplogroup Assignment

2.1.1 Haplogrep

Weissensteiner, H. et al. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic acids research 44(W1), W58-63

- assigns haplogroups based on phylotree and uses a generic rule-based system for immediate quality control
- vcf input

2.1.2 Phy-Mer

Navarro-Gomez, D et al (2014). Phy-Mer: a novel alignment-free and reference-independent mitochondrial haplogroup classifier. Bioinformatics (Oxford, England) 31(8), 1310-2

- novel mitochondrial genome haplogroup-defining algorithm using a k-mer approach by decomposes a mitochondrial sequence into a set of all possible k-mers, which are then compared against each of the k-mer sets of all haplogroups
- input a NGS data (.bam, .cram)

2.2 Estimating mtDNAcn

Mitochondrial DNA Copy Number estimation

- mtDNA-CN can be estimated as the ratio of the average mitochodnrial DNA coverage by the average autosomal DNA coverage
  - mtDNA-CN = (mtDNA average coverage / autosomal DNA average coverage) * 2

2.2.1 fastMitoCalc

Qian, Y., et al. (2017). **fastMitoCalc: an ultra-fast program to estimate mitochondrial DNA copy number from whole-genome sequences.** Bioinformatics 33(9), 1399-1401.

- uses a randomly selected small subset (0.1%) of the nuclear genome to estimate autosomal DNA coverage accurately for estimation of the mtDNA-CN.

2.2.2 Mosedepth

Pedersen, B., Quinlan, A. (2017). **Mosdepth: quick coverage calculation for genomes and exomes** Bioinformatics 34(5), 867-868.

- Mosdepth uses a simple algorithm that is computationally efficient enableing it to quickly calculating genome-wide sequencing coverage. Not specifically designed for estimating mtDNA-CN, but provides coverage estimates of the autosome and mitochondrial genome.

2.3 Cohorts

Accelerating Medicine Partnership in Alzheimer's Disease (AMP-AD)

Whole genome sequencing data was obtained from three cohorts using AMP-AD knowledge portal.

- ROSMAP
- Mayo
- MSBB

Chapter 3

ROSMAP

The samples that we have profiled come from two prospective studies of aging-The Religious order Study (ROS) and the Memory and Aging Project (MAP)-that recruit older individuals without known dementia and include (1) detailed cognitive, neuroimaging and other ante-mortem phenotyping and (2) an autopsy at the time of death that includes a structured neuropathologic examination. A subset of the ROSMAP samples (n=1200 for 1179 unique deceased participants) underwent whole genome sequencing, with DNA coming from brain tissue (n=806), whole blood (n=389) or lymphocytes transformed with EBV virus (n=5) (Jager et al. (2018)).

Data Dictionaries for ROSMAP can be found at:

- AMP-AD
- RADC

```
rosmap.wgsqc <- read_csv("data/AMPAD_extra/rosmap/WGS_sample_QC_info.csv", guess_max = 10000)
rosmap.pheno <- readxl::read_xlsx("data/AMPAD_extra/rosmap/dataset_641_basic_04-29-2020.xlsx") %>
  mutate(projid = as.numeric(projid))
rosmap.raw <- read_csv('data/AMPAD_extra/rosmap/ROSMAP_Clinical_2019-05_v3.csv') %>%
  select(projid, race, spanish, cts_mmse30_lv, educ) %>%
  left_join(rosmap.pheno, by = 'projid') %>%
  left_join(rosmap.wgsqc, by = 'projid') %>%
  filter(!is.na(WGS_id))

mosdepth <- read_tsv('data/mosdepth/mosdepth_all.txt')
haplogrep <- read_tsv('data/haplogrep/haplogrep_all.txt')

rosmap <- rosmap.raw %>%
  filter(QC == "Pass") %>%
  left_join(select(haplogrep, -study), by = c('WGS_id' = 'SampleID')) %>%
  left_join(select(mosdepth, -study), by = c('WGS_id' = 'SampleID')) %>%
```

```
    mutate(race = as_factor(race),
        race = fct_recode(race,'W' = '1', 'B' = '2'),
        z_mtdnacn = scale(mtcn_avg, center = TRUE, scale = TRUE)[,1],
        spanish = as_factor(spanish),
        spanish = fct_recode(spanish,'Yes' = '1', 'No' = '2'),
        organ = recode(Source.Tissue.Type, 'Blood' = 'blood', 'Blood-PBMC' = 'blood',
                       'Blood-Cerebellum' = 'brain',  'Brain-Anterior Caudate' = 'brai
                       'Brain-Cerebellum' = 'brain', 'Brain-DLPFC' = 'brain',
                       'Brain-Frontal Cortex (BA unknown)' = 'brain',
                       'Brain-Frontal Pole (BA10-12,32)' = 'brain',
                       'Brain-Occipital Association Cortex (BA18,19)' = 'brain',
                       'Brain-PCC' = 'brain', 'Brain-Posterior Cingulate Cortex' = 'br
                       'Brain-region unknown' = 'brain',
                       'lymphocytes _transformed _with EBV virus'= 'lymphocytes'),
        organ = as_factor(organ),
        ad_reagan = fct_recode(niareagansc, "1" = "1", "1" = "2", "0" = "3", "0" = "4"
        apoe4 = recode(apoe_genotype, '22' = 'e4-', '23' = 'e4-', '33' = 'e4-', '24' =
        aod_cat = cut(age_death, c(50, 60, 70, 80, 90, Inf), c('50-59', '60-69', '70-7
        aod_cat = ordered(aod_cat, levels = c('50-59', '60-69', '70-79', '80-89', '90+
        msex = as.factor(msex),
        msex = fct_recode(msex, 'M' = '1', 'F' = '0'),
        cogdx = factor(cogdx),
        dcfdx_lv = factor(dcfdx_lv),
        apoe_genotype = as.factor(apoe_genotype),
        apoe4 = as.factor(apoe4),
        study = as.factor(study),
        braaksc = ordered(braaksc, levels = c('0', '1', '2', '3', '4', '5', '6')),
        ceradsc = ordered(ceradsc, levels = c('4', '3', '2', '1')),
        dlbdx = as.factor(dlbdx),
        ci_num2_mct = as.factor(ci_num2_mct),
        ci_num2_gct = as.factor(ci_num2_gct),
        cvda_4gp2 = as.factor(cvda_4gp2),
        caa_4gp = as.factor(caa_4gp),
        arteriol_scler = as.factor(arteriol_scler),
         hspath_typ = as.factor(hspath_typ),
         tdp_st4 = as.factor(tdp_st4),
        niareagansc = ordered(niareagansc, levels = c('4', '3', '2', '1')),
        CDR = cut(cts_mmse30_lv, breaks = c(-Inf, 11, 21, 26, 30, Inf), labels = c(3, 1
  filter(!is.na(study))

saveRDS(rosmap, 'output/rosmap.RData')

df <- rosmap %>%
    select(study, age_bl, msex, educ, apoe_genotype, cogdx, age_first_ad_dx, Source.Tis
```

3.1 Demographics

Demographic variables avaliable in ROSMAP are show in Table 3.1.

Data Summary

variables

definitions

types

missing_percent

unique_count

study

Study

factor

0

2

race

Racial group

factor

0

2

spanish

Spanish ethnicity

factor

0

2

msex

Sex

factor

0

2

educ

Education

numeric

0

25

age_bl

Age at baseline

numeric

0

1100

age_death

Age at death

numeric

0

1071

Descriptive statistics of numerica varibles are presented in Table 3.1.

Variable type: Numeric

col_name

min

q1

median

mean

q3

max

sd

pcnt_na

educ

5.00

14.00

16.00

16.38

19.00

30.00

3.60

0

age_bl

63.02

76.17

81.30

80.85

85.48

102.15

6.89

0

age_death

65.99

84.78

89.18

88.93

93.39

108.28

6.54

0

Frequency and proportions of categorical varibles are presented in Table 3.1.

Variable type: Factor

col_name

level

prop

cnt

msex

F

0.66

779

msex

M

0.34

400

race

W

1.00

1178

race

B

0.00

1

spanish

No

1.00

1178

spanish

Yes

0.00

1

study

MAP

0.51

597

study

ROS

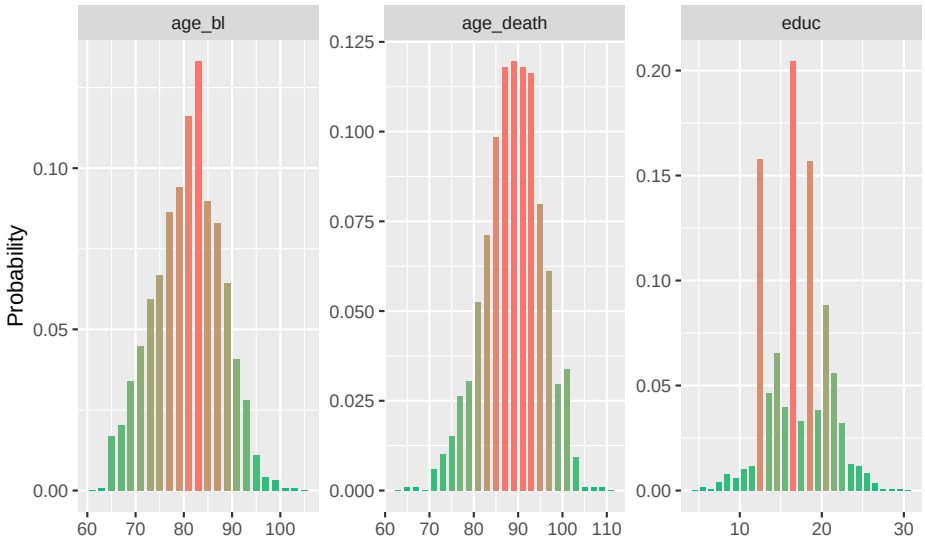
0.49

582

3.1.1 Plots

```
demo_n %>% show_plot()
```

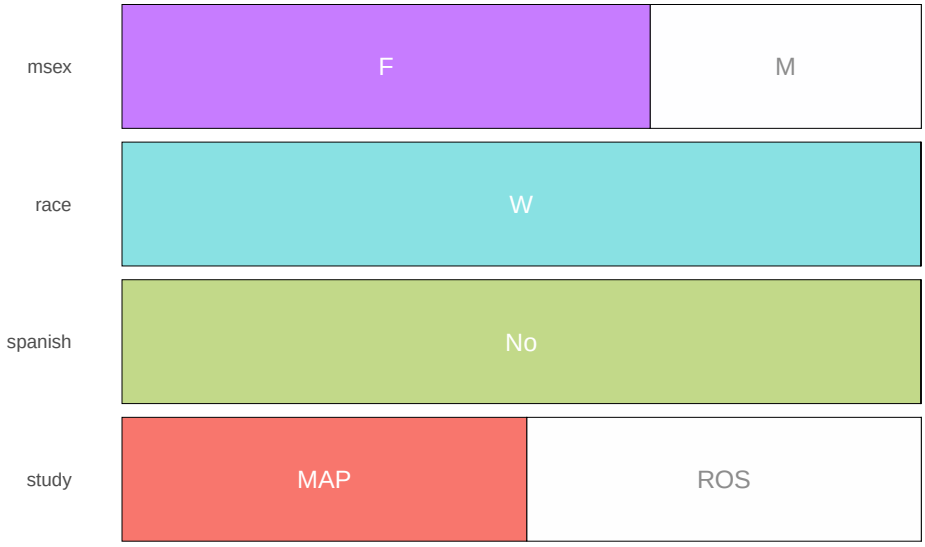
Histograms of numeric columns in df::rosmap




```
demo_c %>% show_plot(high_cardinality = 5)
```

Frequency of categorical levels in df::rosmap

Gray segments are missing values



3.2 Genetics

Data Summary

variables

definitions

types

missing_percent

unique_count

Source.Tissue.Type

Source tissue for DNA

character

0.00

14

organ

collapsed source tissue into organ

factor

0.00

3

apoe_genotype

APOE genotypes

factor

0.76

7

apoe4

APOE e4 carriers

factor

0.76

3

Variable type: Factor

col_name

level

prop

cnt

apoe_genotype

33

0.61

714

apoe_genotype

34

0.22

264

apoe_genotype

23

0.12

145

apoe_genotype

24

0.02

22

apoe_genotype

44

0.02

18

apoe_genotype

NA

0.01

9

apoe_genotype

22

0.01

7

apoe4

e4-
0.73
866
apoe4
e4+
0.26
304
apoe4
NA
0.01
9
organ
brain
0.68
796
organ
blood
0.32
378
organ
lymphocytes
0.00
5
Source.Tissue.Type
Brain-DLPFC
0.39
460
Source.Tissue.Type
Whole Blood
0.30
355

Source.Tissue.Type

Brain-Cerebellum

0.22

256

Source.Tissue.Type

Brain-Posterior Cingulate Cortex

0.06

67

Source.Tissue.Type

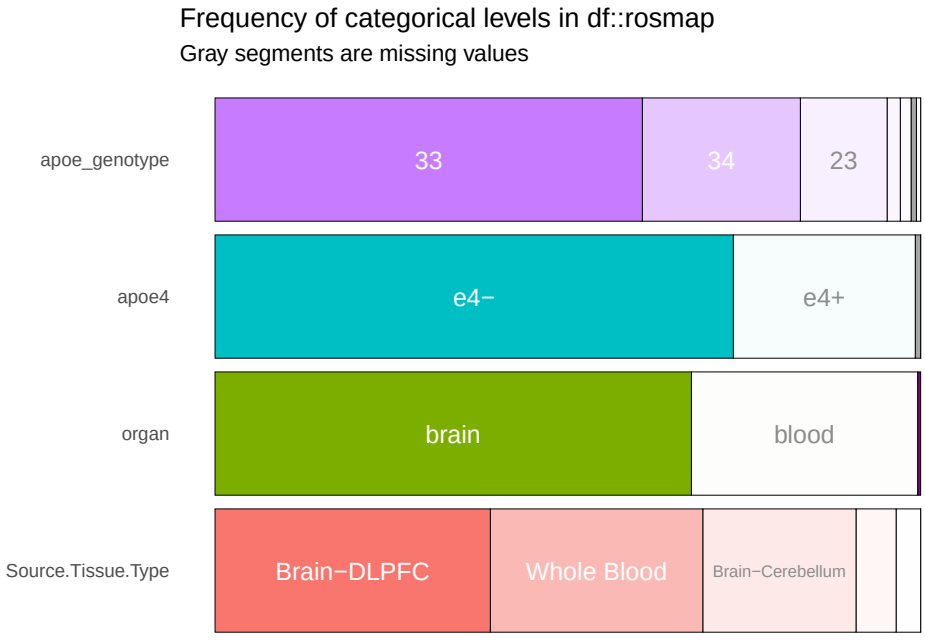
Other

0.03

41

3.2.1 Plots

```
genetic_c %>% show_plot(high_cardinality = 5)
```



3.3 Mitochondria

Variable type: Numeric

col_name

min

q1

median

mean

q3

max

sd

pcnt_na

autosomal_coverage

26.90

33.74

36.41

36.60

39.19

60.26

4.46

1.36

mt_coverage

580.79

10103.45

25751.14

31506.63

53300.38

88911.89

23908.87

1.36

mtcn_avg

41.37

551.39

1410.79

1733.19

2946.14

4988.99

1308.36

1.36

Quality

0.50

0.92

0.95

0.94

0.97

1.01

0.05

1.36

Variable type: Factor

col_name

level

prop

cnt

Haplogroup

Other

0.88

1039

Haplogroup

T2b

0.01

17

Haplogroup

H
0.01
16
Haplogroup
T1a1
0.01
16
Haplogroup
V
0.01
16
Haplogroup
NA
0.01
16
Haplogroup
H1e1a
0.01
11
Haplogroup
H1a
0.01
10
Haplogroup
H1c
0.01
10
Haplogroup
H3
0.01
10

Haplogroup
H5a1
0.01
9
Haplogroup
U5a1a1
0.01
9
macro
H
0.45
534
macro
U
0.16
186
macro
T
0.09
108
macro
J
0.09
103
macro
K
0.06
76
macro
V
0.05

54

macro

I

0.03

37

macro

X

0.02

20

macro

W

0.02

18

macro

NA

0.01

16

macro

Other

0.01

15

macro

N

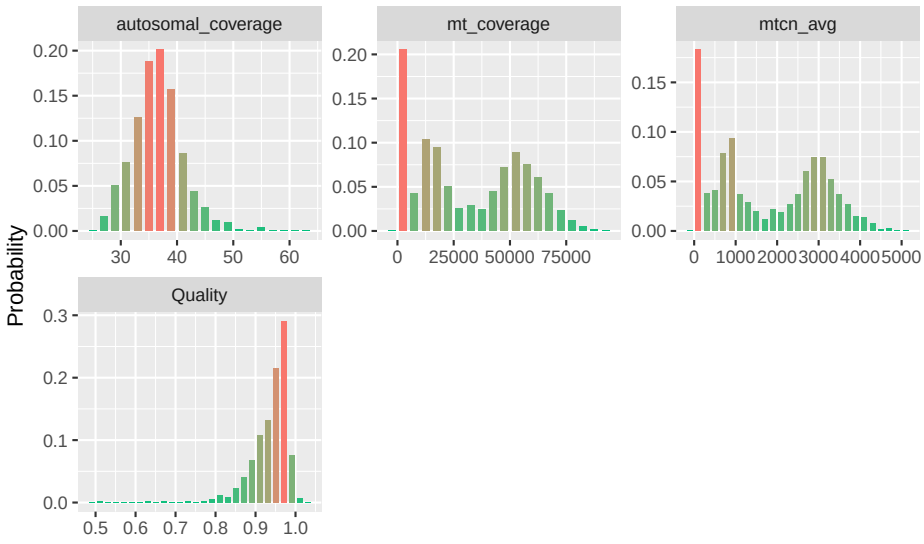
0.01

12

3.3.1 Plots

```
mt_n %>% show_plot()
```

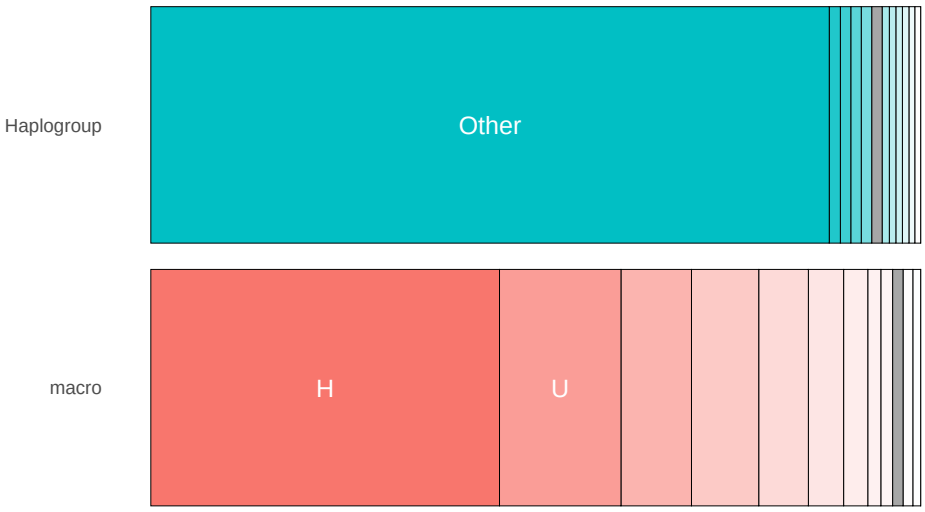
Histograms of numeric columns in df::rosmap




```
mt_c %>% show_plot(high_cardinality = 5)
```

Frequency of categorical levels in df::rosmap

Gray segments are missing values



3.4 Clinical Diagnosis

- Clinical cognitive diagnosis summary: `cogdx` Physician's overall cognitive diagnostic category

  - 1 = NCI: No cognitive impairment (No impaired domains)
  - 2 = MCI: Mild cognitive impairment (One impaired domain) and NO other cause of CI
  - 3 = MCI: Mild cognitive impairment (One impaired domain) AND another cause of CI
  - 4 = AD: Alzheimer's dementia and NO other cause of CI (NINCDS PROB AD)
  - 5 = AD: Alzheimer's dementia AND another cause of CI (NINCDS POSS AD)
  - 6 = Other dementia: Other primary cause of dementia
- Age at first Alzheimer's dementia dx: `age_first_ad_dx` Age at cycle where first Alzheimer's dementia diagnosis was given
- Final consensus cognitive diagnosis: `dcfdx_lv` Clinical consensus diagnosis of cognitive status at time of death - same coding as `cogdx`

Data Summary

variables

definitions

types

missing_percent

unique_count

cogdx

Physician's overall cognitive diagnostic category

factor

0.00

6

age_first_ad_dx

Age at cycle where first Alzheimer's dementia diagnosis was given

numeric

65.14

392

dcfdx_lv

Clinical consensus diagnosis of cognitive status at time of death

factor

0.00

6

Variable type: Numeric

col_name

min

q1

median

mean

q3

max

sd

pcnt_na

age_first_ad_dx

68.93

83.21

87.38

87.32

91.36

107.23

6.39

65.14

Variable type: Factor

col_name

level

prop

cnt

cogdx

4

0.37

433

1

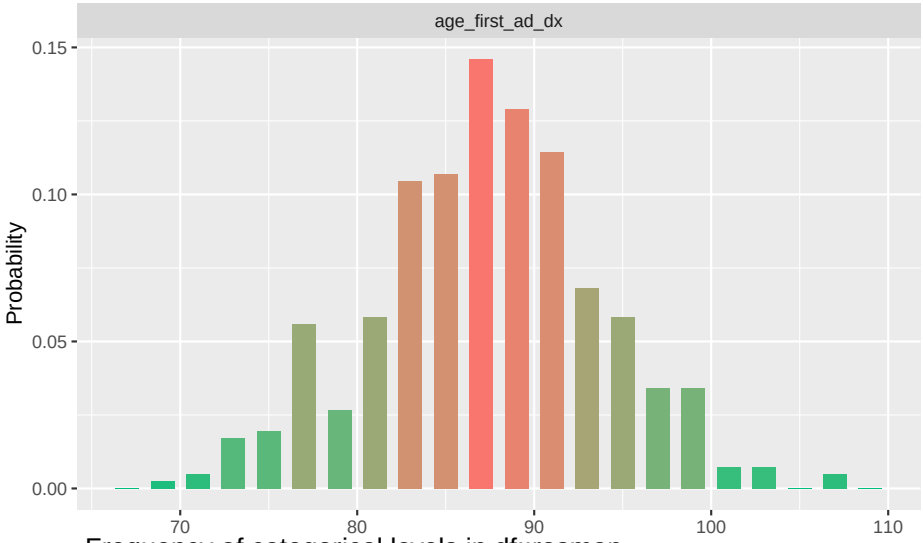
0.32

374

2
0.23
273
5
0.05
58
6
0.02
21
3
0.02
20
dcfdx_lv
4
0.36
421
1
0.32
379
2
0.25
289
5
0.05
60
6
0.02
20
3
0.01
10

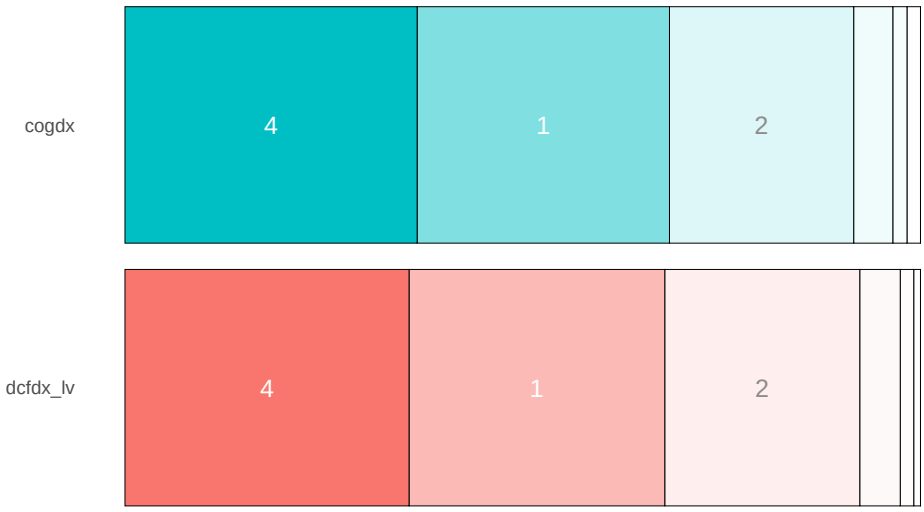
3.4.1 Plots

Histograms of numeric columns in df::rosmap



Frequency of categorical levels in df::rosmap

Gray segments are missing values



3.4.2 Cross-tabs

Characteristic

1

2

3

4

5

6

Total

cogdx

1

359 (30%)

14 (1.2%)

0 (0%)

1 (<0.1%)

0 (0%)

0 (0%)

374 (32%)

2

14 (1.2%)

251 (21%)

5 (0.4%)

2 (0.2%)

0 (0%)

1 (<0.1%)

273 (23%)

3

4 (0.3%)

11 (0.9%)

4 (0.3%)

1 (<0.1%)

0 (0%)

0 (0%)

20 (1.7%)

4
1 (<0.1%)
13 (1.1%)
0 (0%)
391 (33%)
22 (1.9%)
6 (0.5%)
433 (37%)
5
0 (0%)
0 (0%)
0 (0%)
17 (1.4%)
34 (2.9%)
7 (0.6%)
58 (4.9%)
6
1 (<0.1%)
0 (0%)
1 (<0.1%)
9 (0.8%)
4 (0.3%)
6 (0.5%)
21 (1.8%)
Total
379 (32%)
289 (25%)
10 (0.8%)
421 (36%)
60 (5.1%)
20 (1.7%)

1179 (100%)

3.5 Pathology

Pathology: post-mortem neuropathologic evaluation

- Alzheimer's disease
  - NIA-Reagan diagnosis of AD: `niareaganscmodified` NIA-Reagan diagnosis of Alzheimer's disease is based on consensus recommendations for postmortem diagnosis of Alzheimer's disease. The criteria rely on both neurofibrillary tangles (Braak) and neuritic plaques (CERAD).
    - 1 = High; 2 = Intermediate; 3 = Low; 4 = No AD
  - Dichomtomized NIA-Reagan: `ad_reagan`
  - CERAD score: `ceradsc` CERAD score is a semiquantitative measure of neuritic plaques. A CERAD neuropathologic diagnosis of AD required moderate (probable AD) or frequent neuritic plaques (definite AD) in one or more neocortical regions.
    - 1 = Definite -> frequent (C3); 2 = Probable -> moderate (C2); 3 = Possible -> Sparse (C1); 4 = No AD -> None (C0)
  - Braak stage: `braaksc` Braak Stage is a semiquantitative measure of severity of neurofibrillary tangle (NFT) pathology.
    - 0 = 0; 1 = I (entorhinal); 2 = II (entorhinal); 3 = III (limbic); 4 = IV (limbic); 5 = V (neocortical); 6 = VI (neocortical)
  - Global AD pathology burden: `gpath` Global AD pathology burden is a quantitative summary of AD pathology derived from counts of three AD pathologies: neuritic plaques (n), diffuse plaques (d), and neurofibrillary tangles (nft)
- Beta-Amyloid
  - amyloid: `amyloid` Overall amyloid level - Mean of 8 brain regions
  - plaq_d: `plaq_d` Diffuse plaque summary based on 5 regions
  - plaq_n: `plaq_n` Neuritic plaque summary based on 5 regions
- PHF tau Tangles
  - Tangle: `tangles` Tangle density - Mean of 8 brain regions
  - NFT burden: `nft` Neurofibrillary tangle summary based on 5 regions
- Lewy Body disease: `dlbdx` Pathologic diagnosis of Lewy body diseases - 4 stages
  - 0 = Not present; 1 = nigral-predominant; 2 = limbic-type; 3 = neocortical-type
- Vascular
  - gross infarcts: `ci_num_gct` Cerebral Infarctions - Binary - Gross-Chronic-Any Location
  - micro infarcts: `ci_num2_mct` Cerebral Infarctions - Binary - Micro-Chronic-Any Location
  - Cerebral atherosclerosis: `cvda_4gp2` Cerebral Atherosclerosis Rating

        * 0 = None; 1 = Mild; 2 = Moderate; 3 = Severe
    - Cerebral amyloid angiopathy: `caa_4gp` Cerebral amyloid angiopathy
        * 0 = None; 1 = Mild; 2 = Moderate; 3 = Severe
    - Arteriolosclerosis: `arteriol_scler` Arteriolosclerosis
        * 0 = None; 1 = Mild; 2 = Moderate; 3 = Severe
- Hippocampal sclerosis (Typical): `hspath_typ` Definite presence of typical hippocampal sclerosis
- TDP-43 stage: `tdp_st4` TDP-43 pathology from 8 regions
    - 0 = None; 1 = Amygdala; 2 = Amygdala + Limbic; 3 = Amygdala + Limbic + Neocortical

Data Summary

variables

definitions

types

missing_percent

unique_count

niareagansc

NA

ordered

0.00

4

ceradsc

NA

ordered

0.00

4

braaksc

NA

ordered

0.00

7

gpath

NA

numeric

0.00

1121

amyloid

NA

numeric

0.68

990

plaq_d

NA

numeric

0.00

910

plaq_n

NA

numeric

0.00

879

tangles

NA

numeric

1.02

1151

nft

NA

numeric

0.00

979

dlbdx

NA

factor

3.31

5
ci_num2_gct
NA
factor
0.00
2
ci_num2_mct
NA
factor
0.00
2
cvda_4gp2
NA
factor
0.59
5
caa_4gp
NA
factor
3.05
5
arteriol_scler
NA
factor
0.68
5
hspath_typ
NA
factor
0.85
3

tdp_st4

NA

factor

8.40

5

Variable type: Numeric

col_name

min

q1

median

mean

q3

max

sd

pcnt_na

gpath

0

0.19

0.66

0.76

1.17

2.95

0.63

0.00

amyloid

0

0.62

3.05

4.26

6.67

22.94

4.23

0.68

plaq_d

0

0.08

0.58

0.79

1.18

4.91

0.82

0.00

plaq_n

0

0.06

0.71

0.84

1.33

5.01

0.83

0.00

tangles

0

1.49

4.01

6.65

8.56

61.01

7.76

1.02

nft

0

0.13

0.37

0.65

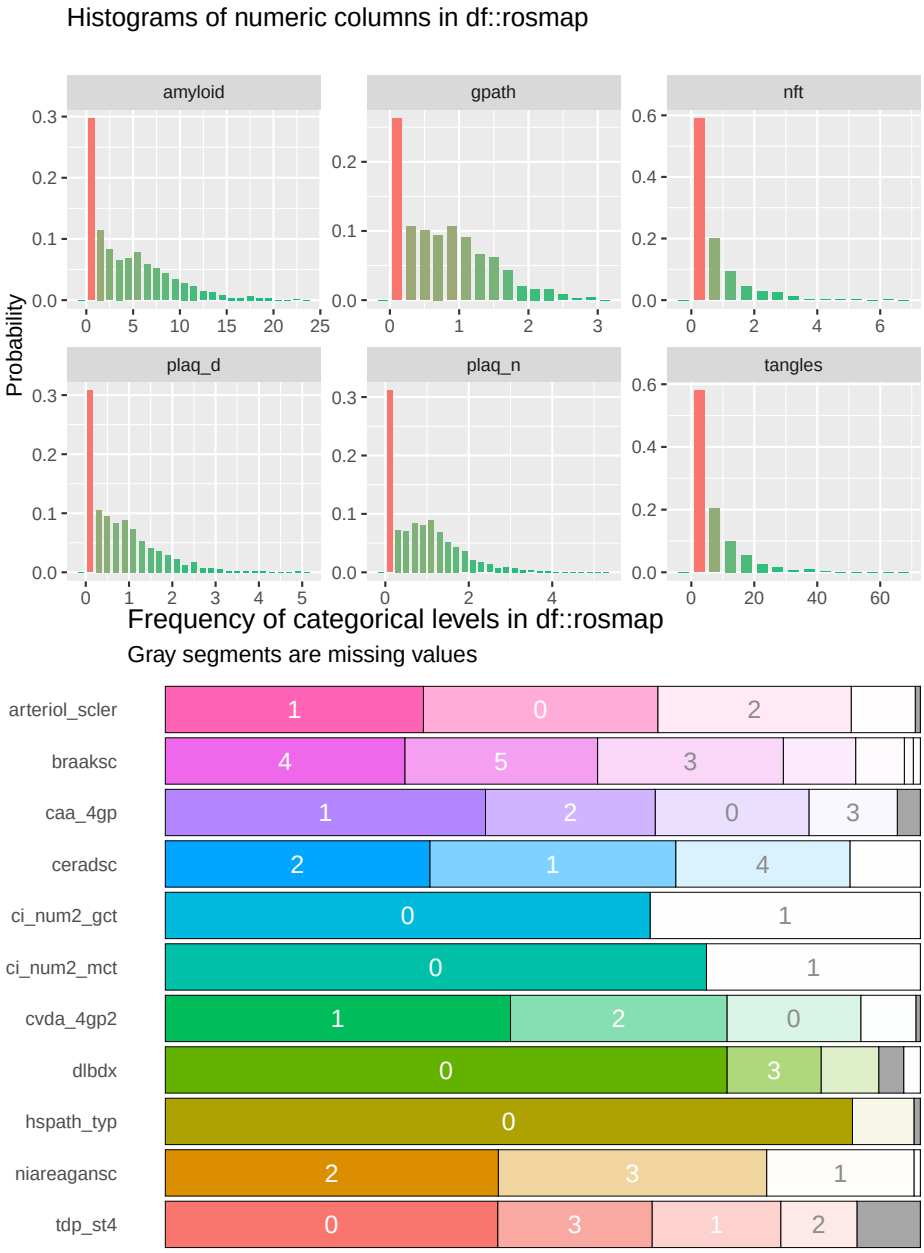
0.86

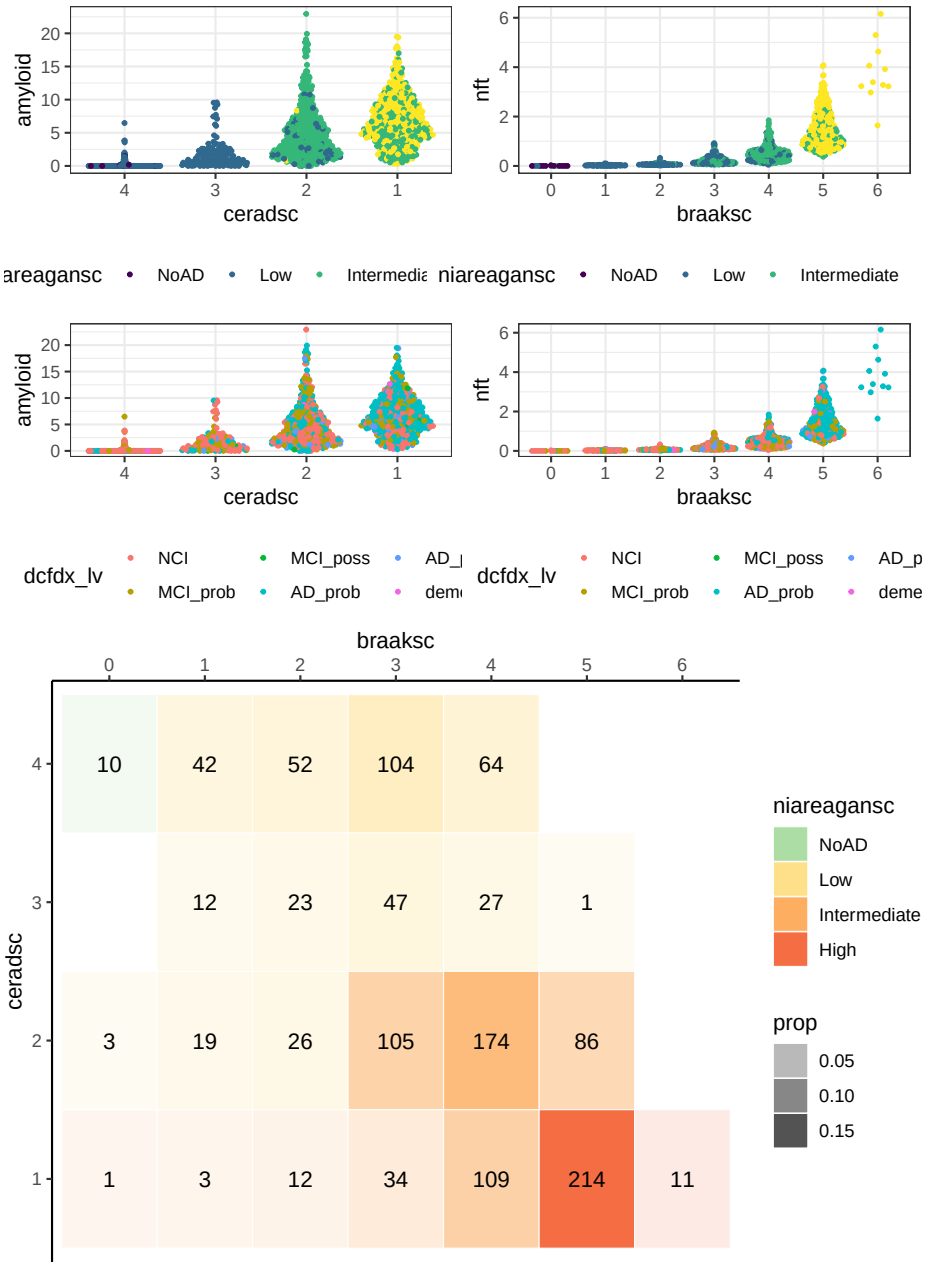
6.16

0.76

0.00

3.5.1 Plots


Histograms of numeric columns in df::rosmap
amyloid
gpath
nft
plaq_d
plaq_n
tangles
Probability
Frequency of categorical levels in df::rosmap
Gray segments are missing values
arteriol_scler
braaksc
caa_4gp
ceradsc
ci_num2_gct
ci_num2_mct
cvda_4gp2
dlbdx
hspath_typ
niareagansc
tdp_st4



3.5.2 Cross-Tabs

Characteristic

0

1

2

3

4

5

6

Total

ceradsc

4

10 (0.8%)

42 (3.6%)

52 (4.4%)

104 (8.8%)

64 (5.4%)

0 (0%)

0 (0%)

272 (23%)

3

0 (0%)

12 (1.0%)

23 (2.0%)

47 (4.0%)

27 (2.3%)

1 (<0.1%)

0 (0%)

110 (9.3%)

2

3 (0.3%)

19 (1.6%)

26 (2.2%)

105 (8.9%)

174 (15%)

86 (7.3%)

0 (0%)

413 (35%)

1

1 (<0.1%)

3 (0.3%)

12 (1.0%)

34 (2.9%)

109 (9.2%)

214 (18%)

11 (0.9%)

384 (33%)

Total

14 (1.2%)

76 (6.4%)

113 (9.6%)

290 (25%)

374 (32%)

301 (26%)

11 (0.9%)

1179 (100%)

Characteristic

4

3

2

1

Total

cogdx

1

6 (1.6%)

208 (56%)
144 (39%)
16 (4.3%)
374 (100%)
2
4 (1.5%)
105 (38%)
133 (49%)
31 (11%)
273 (100%)
3
0 (0%)
11 (55%)
7 (35%)
2 (10%)
20 (100%)
4
0 (0%)
60 (14%)
206 (48%)
167 (39%)
433 (100%)
5
0 (0%)
25 (43%)
22 (38%)
11 (19%)
58 (100%)
6
0 (0%)
10 (48%)

8 (38%)

3 (14%)

21 (100%)

Total

10 (0.8%)

419 (36%)

520 (44%)

230 (20%)

1179 (100%)

ceradsc

braaksc

NoAD

Low

Intermediate

High

C0

B0

10

NA

NA

NA

B1

NA

94

NA

NA

B2

NA

168

NA

NA

C1
B1
NA
35
NA
NA
B2
NA
73
1
NA
B3
NA
1
NA
NA
C2
B0
NA
3
NA
NA
B1
NA
44
1
NA
B2
NA
1
278

NA

B3

NA

NA

82

4

C3

B0

NA

NA

1

NA

B1

NA

NA

15

NA

B2

NA

NA

142

1

B3

NA

NA

NA

225

3.6 Cognition

mmse: Mini Mental State Examination is a widely used, 30 item, standardized screening measure of dementia severity. The MMSE can be used as a surrogate measure for the CDR for the staging of dementia in AD Perneczky et al 2006.

MMSE	CDR
30	0 (No)
26-29	0.5 (Questionable)
21-25	1 (Mild)
11-20	2 (Moderate)
0-10	3 (Severe)

```
# A tibble: 2 x 6
  variables     types   missing_count missing_percent unique_count unique_rate
  <chr>         <chr>           <int>           <dbl>        <int>       <dbl>
1 cts_mmse30_lv numeric             2           0.170           84     0.0712
2 CDR           factor              2           0.170            6     0.00509
```

Variable type: Numeric

col_name

min

q1

median

mean

q3

max

sd

pcnt_na

cts_mmse30_lv

0

15.56

25

20.84

28

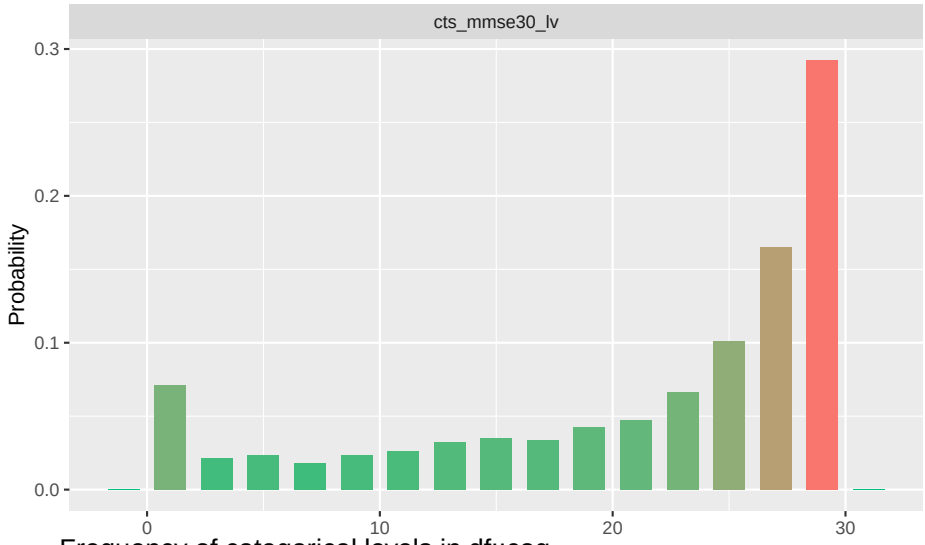
30

9.2

0.17

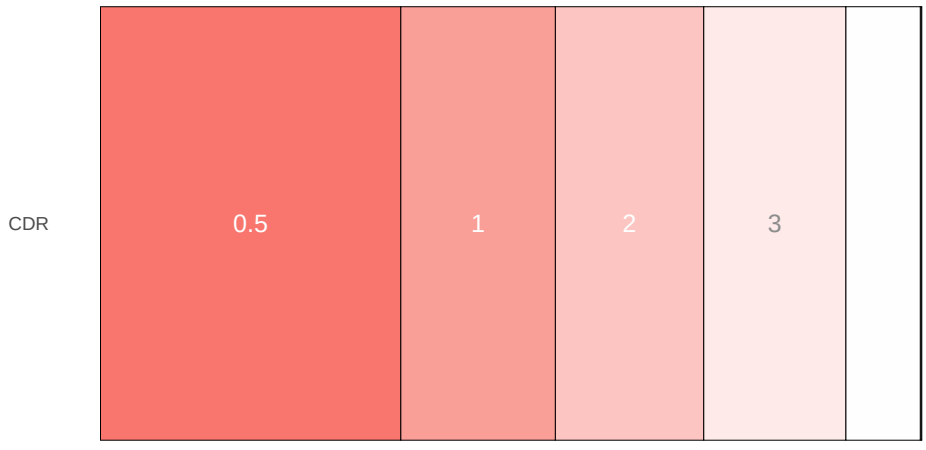
3.6.1 Plots

Histograms of numeric columns in df::cog



Frequency of categorical levels in df::cog

Gray segments are missing values



Chapter 4

MSBB

Samples come from 364 postmortem control, mild cognitive impaired (MCI) and AD brains with rich clinical and pathophysiological data from the Mount Sinai/JJ Peters VA Medical Center Brain Bank (MSBB–Mount Sinai NIH Neurobiobank) cohort. The majority (301) of the samples were of European ancestry, while 36 were African American, 25 were Latino, one was Asian, and one was unknown for race. Neuropathological assessments were performed using CERAD scores and Braak Staging. The CDR scale was conducted for assessment of dementia and cognitive status (Wang et al. (2018)).

Clinical Code Book: Synapse

```
## MSBB
msbb.raw1 <- read_tsv('data/AMPAD/msbb/WGS_Metadata.txt')
msbb.raw2 <- read_tsv('data/AMPAD_extra/msbb.wgs.meta.tsv')
msbb.path <- readxl::read_xlsx('data/AMPAD_extra/msbb/TempAmp_AD-Shea-2-2020.xlsx', sheet = 2) %>
  mutate(id = as.character(id))
msbb.raw <- msbb.raw1 %>%
  select(WGS, individualIdentifier, PMI, RACE, CDR, SEX, NP.1, PlaqueMean, bbscore) %>%
  left_join(select(msbb.raw2, Libid, AOD, APOE_inferred), by = c('WGS' = 'Libid')) %>%
  mutate(SubNum = str_extract(gsub("(?<![0-9])0+", "", individualIdentifier, perl = TRUE), '[:dig
  left_join(msbb.path, by = c('SubNum' = 'id')) %>%
  mutate(study = 'MSBB',
         WGS = as.character(WGS))

mosdepth <- read_tsv('data/mosdepth/mosdepth_all.txt')
haplogrep <- read_tsv('data/haplogrep/haplogrep_all.txt')
```

```
## Recode NP.1 -> ceradsc to be consistent with ROSMAP
msbb <- msbb.raw %>%
  left_join(haplogrep, by = c('WGS' = 'SampleID')) %>%
  left_join(mosdepth, by = c('WGS' = 'SampleID')) %>%
```

```
  mutate(id = paste0('MSBB', WGS)) %>%
  mutate(APOE_inferred = recode(APOE_inferred, 'e2/e2' = '22', 'e2/e3' = '23', 'e3/e3'
         apoe4 = recode(APOE_inferred, '22' = 'e4-', '23' = 'e4-', '33' = 'e4-', '24' =
         SEX = as.factor(SEX),
         SourceTissue = 'prefrontal cortex',
         APOE_inferred = as.factor(APOE_inferred),
         apoe4 = as.factor(apoe4),
         RACE = as.factor(RACE),
         z_mtdnacn = scale(mtcn_avg, center = TRUE, scale = TRUE)[,1],
         cerad = recode(NP.1, '1' = 'Normal', '2' = 'Definite', '3' = 'Probable', '4' =
         cerad = ordered(cerad, levels = c('Normal', 'Possible', 'Probable', 'Definite'
         ceradsc = pmax(HippoPlaquesWCoresValue, EntorPlaquesWCoresValue, MidPlaquesWCo
                        SupPlaquesWCoresValue, InfPlaquesWCoresValue, OcciPlaquesWCores
         ceradsc = ordered(ceradsc),
         ceradsc = fct_recode(ceradsc, '4' = '0', '3' = '1', '2' = '3', '1' = '5'),
         bbscore = ordered(bbscore, levels = c('0', '1', '2', '3', '4', '5', '6')),
         niareagansc = case_when(
           ceradsc == 4 & bbscore == 0 ~ 4,
           ceradsc == 4 & bbscore %in% c(1:6) ~ 3,
           ceradsc == 3 & bbscore %in% c(0:6) ~ 3,
           ceradsc == 2 & bbscore %in% c(0:2) ~ 3,
           ceradsc == 2 & bbscore %in% c(3:6) ~ 2,
           ceradsc == 1 & bbscore %in% c(0:4) ~ 2,
           ceradsc == 1 & bbscore %in% c(5:6) ~ 1,
         ),
         niareagansc = ordered(niareagansc, levels = c('4', '3', '2', '1')),
         ad_reagan = fct_recode(niareagansc, "1" = "1", "1" = "2", "0" = "3", "0" = "4"
         braaksc_B = fct_recode(bbscore, B0 = "0", B1 = "1", B1 = "2", B2 = "3", B2 = "
         ceradsc_C = fct_recode(ceradsc, C0 = "1", C1 = "2", C2 = "3", C3 = "4")) %>%
  mutate_at(vars(HippoPlaquesWCoresValue, EntorPlaquesWCoresValue, MidPlaquesWCoresVal
                        SupPlaquesWCoresValue, InfPlaquesWCoresValue, OcciPlaquesWCores
            ordered) %>%
  select(id, study, sex = SEX, race = RACE, SourceTissue, PMI, NP.1, cerad, niareagans

saveRDS(msbb, 'output/msbb.rds')
```

4.1 Pathology

Amyloid

- `HippoPlaquesWCoresValue`, `EntorPlaquesWCoresValue`, `MidPlaquesWCoresValue`, `SupPlaquesWCoresValue`, `InfPlaquesWCoresValue`, `OcciPlaquesWCoresValue`: Neuritic plaque burden measured in 8 brain regions. (0 = Absent; 1 = Sparese; 3 = Moderate; 5 = Frequent)
  - requested from Haroutunian, Vahram on March 4 2020

- `ceradsc`: semiquantitative estimates of neuritic plaque density modified to be implemented without adjustment for age and clinical diagnosis, as implemented in ROSMAP
- score is derived from the brain region with the greatest number of neuritic plaques
- `PlaqueMean`: Average number of plaques across brain regions

Nurofibilary Tangles

- `braaksc`: Braak Stage is a semiquantitative measure of severity of neurofibrillary tangle (NFT) pathology

Neuropathological Diagnosis

- `cerad/NP.1`: Neuropathology Category as measured by CERAD (1=Normal, 2=Definite AD, 3=probable AD, 4=possible AD)
- `niareaganscc`: modified NIA-Reagan diagnosis of Alzheimer's disease is based on consensus recommendations for postmortem diagnosis of Alzheimer's disease. The criteria rely on both neurofibrillary tangles (Braak) and neuritic plaques (CERAD) and does not account for clinical information.
  - 1 = High; 2 = Intermediate; 3 = Low; 4 = No AD
  - Implemented to match coding from ROSMAP.

- `ad_reagan`: dichotomized NIA-Reagan diagnosis

Variable type: Numeric

col_name

min

q1

median

mean

q3

max

sd

pcnt_na

PlaqueMean

0

2.11

6.14

8.81

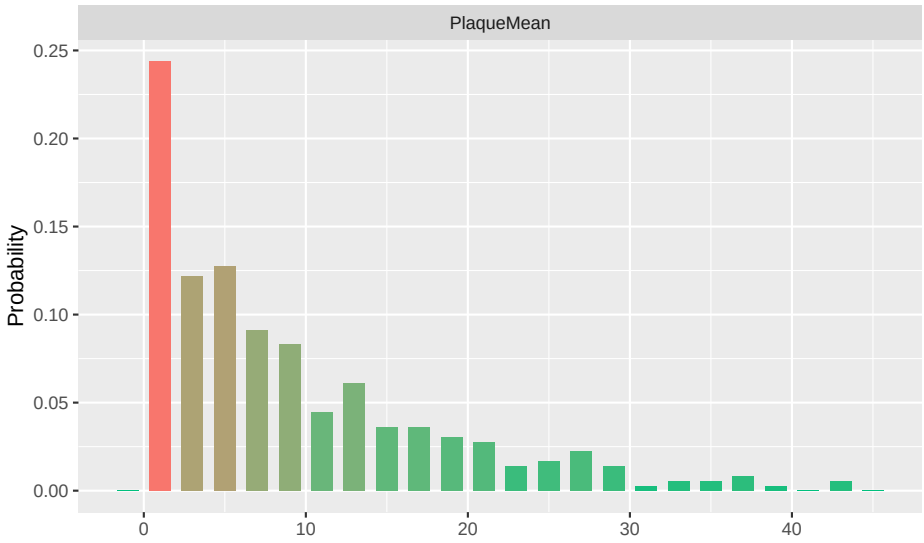
12.38

43.84

8.98

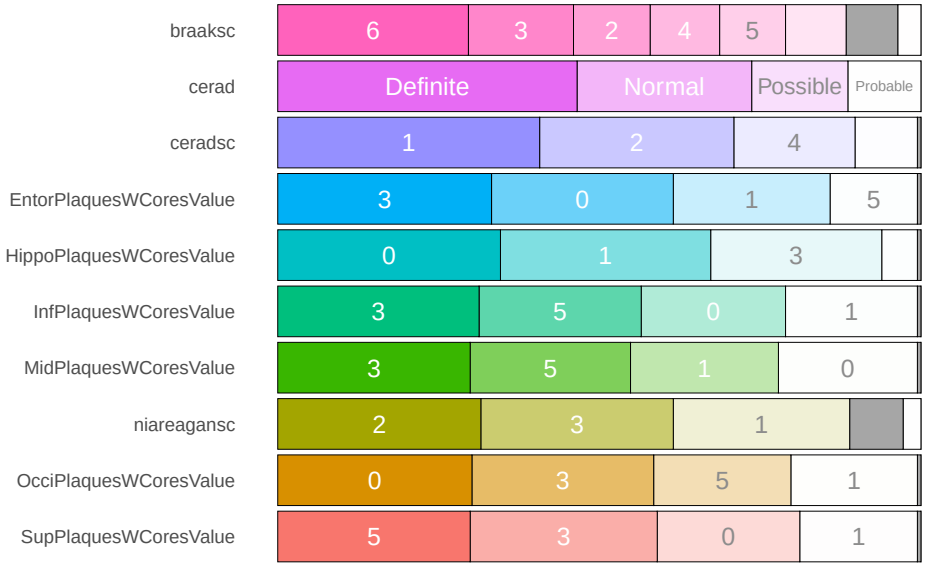
0

Histograms of numeric columns in df::msbb



Frequency of categorical levels in df::msbb

Gray segments are missing values



4.1.1 Cross-tabs

Characteristic

0

1

2

3

4

5

6

Unknown

Total

ceradsc

4

10 (2.8%)

16 (4.4%)

22 (6.1%)

16 (4.4%)

2 (0.6%)

0 (0%)

0 (0%)

2 (0.6%)

68 (19%)

3

2 (0.6%)

2 (0.6%)

8 (2.2%)

10 (2.8%)

3 (0.8%)

1 (0.3%)

0 (0%)

9 (2.5%)
35 (9.7%)
2
1 (0.3%)
15 (4.2%)
10 (2.8%)
20 (5.5%)
11 (3.0%)
18 (5.0%)
26 (7.2%)
8 (2.2%)
109 (30%)
1
0 (0%)
1 (0.3%)
3 (0.8%)
13 (3.6%)
22 (6.1%)
18 (5.0%)
81 (22%)
9 (2.5%)
147 (41%)
Unknown
0 (0%)
0 (0%)
0 (0%)
0 (0%)
1 (0.3%)
0 (0%)
0 (0%)
1 (0.3%)

2 (0.6%)

Total

13 (3.6%)

34 (9.4%)

43 (12%)

59 (16%)

39 (11%)

37 (10%)

107 (30%)

29 (8.0%)

361 (100%)

Characteristic

NoAD

Low

Intermediate

High

Unknown

Total

cerad

Normal

10 (2.8%)

72 (20%)

3 (0.8%)

0 (0%)

13 (3.6%)

98 (27%)

Possible

0 (0%)

26 (7.2%)

28 (7.8%)

0 (0%)

0 (0%)

54 (15%)

Probable

0 (0%)

6 (1.7%)

31 (8.6%)

0 (0%)

4 (1.1%)

41 (11%)

Definite

0 (0%)

4 (1.1%)

52 (14%)

99 (27%)

13 (3.6%)

168 (47%)

Total

10 (2.8%)

108 (30%)

114 (32%)

99 (27%)

30 (8.3%)

361 (100%)

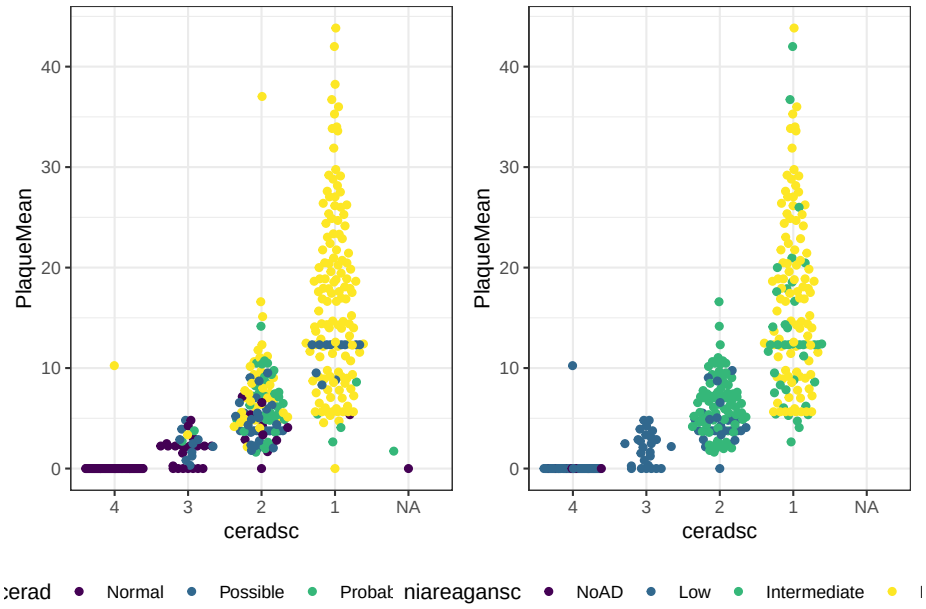


Figure 4.1: Distribution of amyloid by neuropathological diagnosis

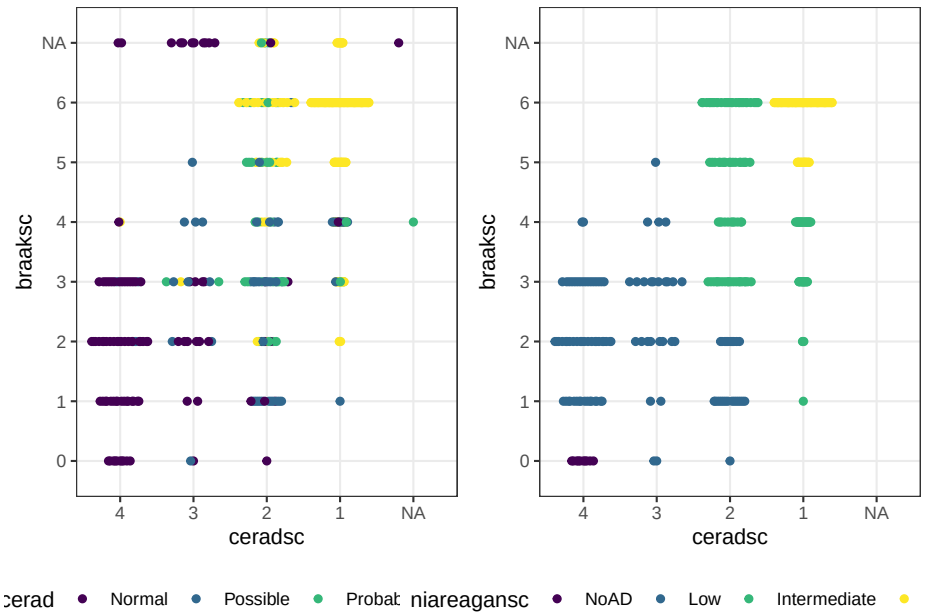
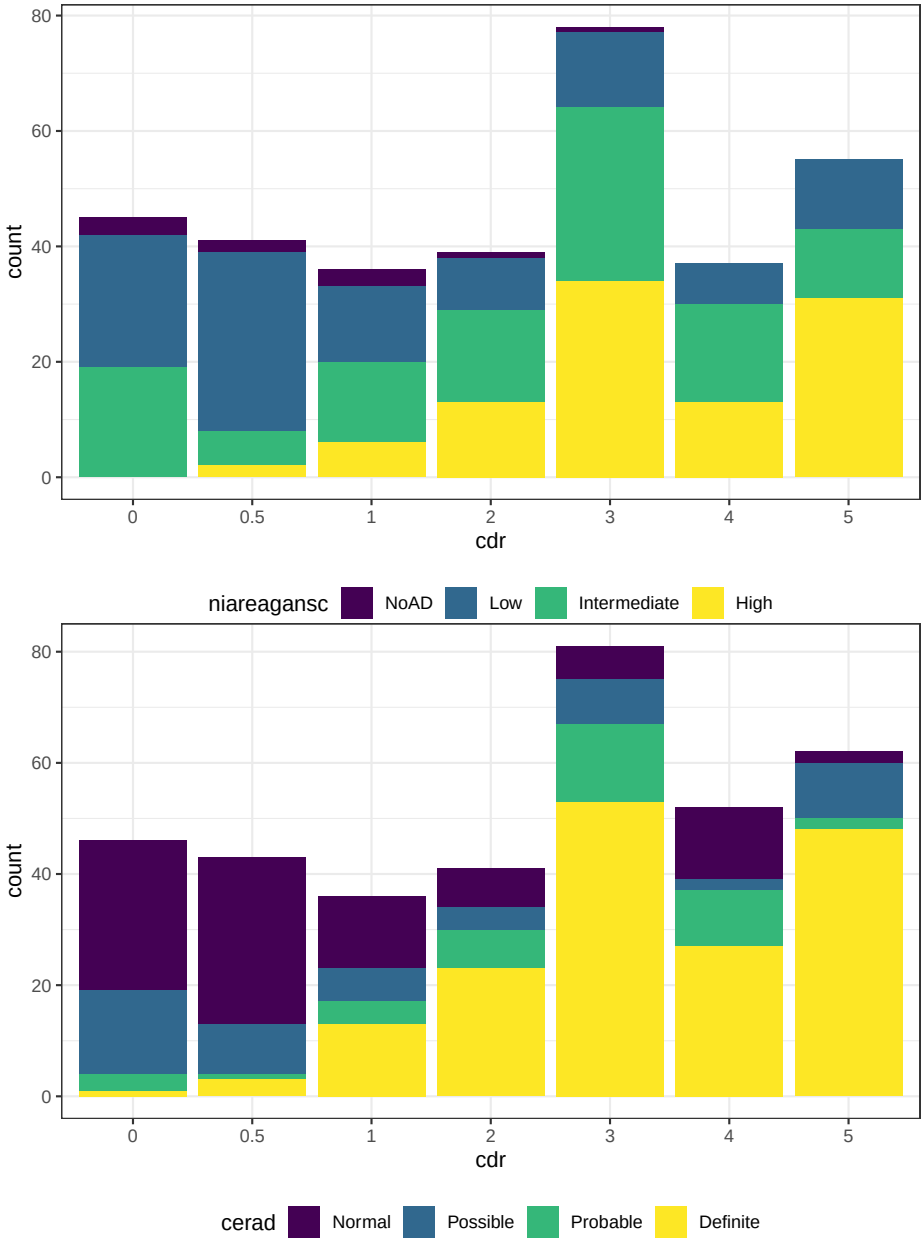


Figure 4.2: Distribution of neuropathological diagnosis by ceradsc and braaksc

4.1.2 Plots



4.2 Other Variables

Variable type: Numeric

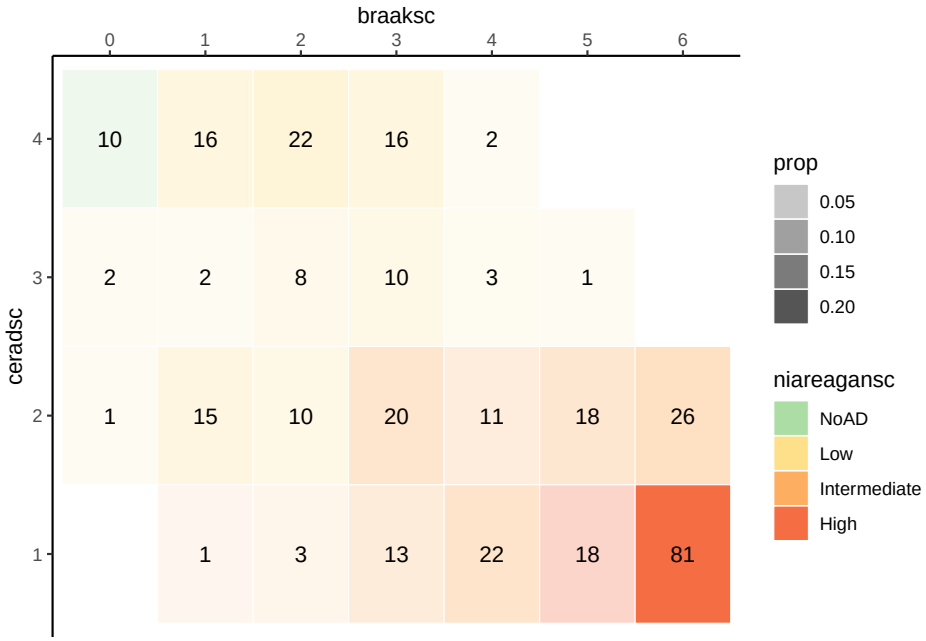


Figure 4.3: Cross-tabs of cerad & braaksc

col_name

min

q1

median

mean

q3

max

sd

pcnt_na

PMI

75.00

220.00

315.00

437.21

537.00

1800.00

325.97

0.00

cdr

0.00

1.00

3.00

2.49

4.00

5.00

1.73

0.00

aod

61.00

79.50

85.00

85.09

92.00

108.00

9.37

0.55

mtcn_avg

734.20

1513.48

1738.74

1726.30

1943.53

3151.59

340.58

0.00

Quality

0.62

0.93

0.96

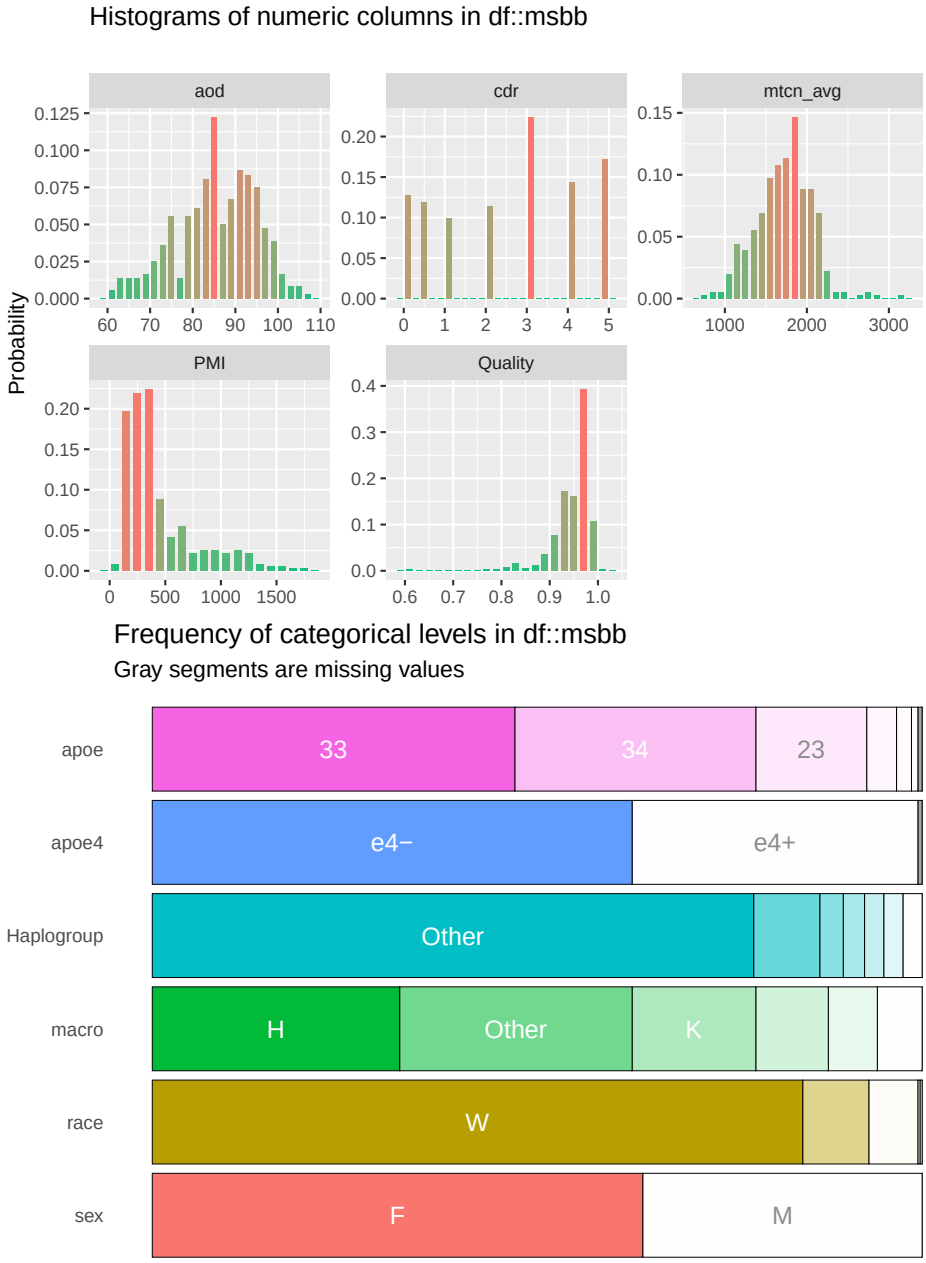
0.95

0.97

1.01

0.04

0.00


Histograms of numeric columns in df::msbb
aod
cdr
mtcn_avg
PMI
Quality
Probability
Frequency of categorical levels in df::msbb
Gray segments are missing values
apoe
33
34
23
apoe4
e4−
e4+
Haplogroup
Other
macro
H
Other
K
race
W
sex
F
M

Chapter 5

MAYO

Allen et al *Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases* Scientific Data 2016

Mayo Clinic Alzheimer's Disease Genetics Studies (MCADGS). Data is provided for the Mayo RNAseq Study, with whole transcriptome data for 275 Cerebellum (CBE) and 276 Temporal cortex (TCX) samples from 312 North American Caucasian subjects with neuropathological diagnosis of AD, progressive supranuclear palsy (PSP), pathologic aging (PA) or elderly controls (CON) without neurodegenerative diseases. Whole genome sequencing was conducted on 349 participants using DNA isolated from either the Temporal cortex (n = 341) or the Cerebellar Cortex (n = 8).

- All ADs had definite diagnosis according to the NINCDS-ADRDA criteria and had Braak NFT stage of IV or greater.
- Control subjects had Braak NFT stage of III or less, CERAD neuritic and cortical plaque densities of 0 (none) or 1 (sparse) and lacked any of the following pathologic diagnoses: AD, Parkinson's disease (PD), DLB, VaD, PSP, motor neuron disease (MND), CBD, Pick's disease (PiD), Huntington's disease (HD), FTLD, hippocampal sclerosis (HipScl) or dementia lacking distinctive histology (DLDH).
- Subjects with PA also lacked the above diagnoses and had Braak NFT stage of III or less, but had CERAD neuritic and cortical plaque densities of 2 or more. None of the PA subjects had a clinical diagnosis of dementia or mild cognitive impairment.

Clinical Code Book: Synapse

Chapter 6

Analysis

```
library(tidyverse)
library(broom)
library(glue)
library(ggbeeswarm)
library(gvlma)
library(inspectdf)
knitr::opts_knit$set(root.dir = '/sc/arion/projects/LOAD/shea/Projects/mtDNAcn')
```

```
rosmap <- readRDS('output/rosmap.RData')
msbb <- readRDS('output/msbb.rds')
```

6.1 ROSMAP

6.1.1 Data Wrangling

- DNA isolated from Brain tissue
  - DLPFC, Posterior Cingulate Cortex, Cerebellum
- European Haplogroups only
  - H, V, J, T, U, K, I, W, X
- For clinical diagnosis, exclude MCI, AD possible, other dementia

```
rosmap_df <- rosmap %>%
  filter(Source.Tissue.Type %in% c("Brain-DLPFC", "Brain-Posterior Cingulate Cortex", "Brain-Cere
  filter(organ == 'brain') %>%
  filter(macro %in% c('H', 'V', 'J', 'T', 'U', 'K', 'I', 'W', 'X')) %>%
  filter(!is.na(macro)) %>%
  mutate(ad_reagan = fct_relevel(ad_reagan, "0", "1"),
         Source.Tissue.Type = fct_inorder(Source.Tissue.Type),
         dx = fct_recode(dcfdx_lv, "0" = "1", NULL = "2", NULL = "3", "1" = "4", NULL = "5", NULL
```

```
  dplyr::select(ad_reagan, niareagansc, dcfdx_lv, dx, cts_mmse30_lv, pmi, study, age_d
```

6.1.2 NIA-Reagan Diagnosis

```
rosmap_path_res <- glm(ad_reagan ~ z_mtdnacn + macro + age_death + msex + apoe4 + study
                       family = "binomial", data = rosmap_df)
```

Assocation of mtDNA with MMSE in ROSMAP

term

estimate

std.error

statistic

p.value

(Intercept)

-5.864

1.256

-4.670

3.0e-06 ***

z_mtdnacn

-0.458

0.197

-2.320

0.02 *

macroI

-0.031

0.455

-0.068

0.946

macroJ

0.697

0.337

2.071

0.038 *

macroK

0.111

0.327

0.340

0.734

macroT

0.153

0.289

0.529

0.597

macroU

0.015

0.238

0.064

0.949

macroV

0.344

0.405

0.850

0.396

macroW

-0.087

0.634

-0.138

0.89

macroX

0.368

0.717

0.514

0.608

age_death

0.076

0.013

5.624

1.9e-08 ***

msexM

-0.330

0.175

-1.887

0.059 .

apoe4e4+

1.416

0.223

6.357

2.1e-10 ***

studyROS

0.050

0.171

0.294

0.768

Source.Tissue.TypeBrain-Cerebellum

-1.013

0.348

-2.909

0.004 **

Source.Tissue.TypeBrain-Posterior Cingulate Cortex

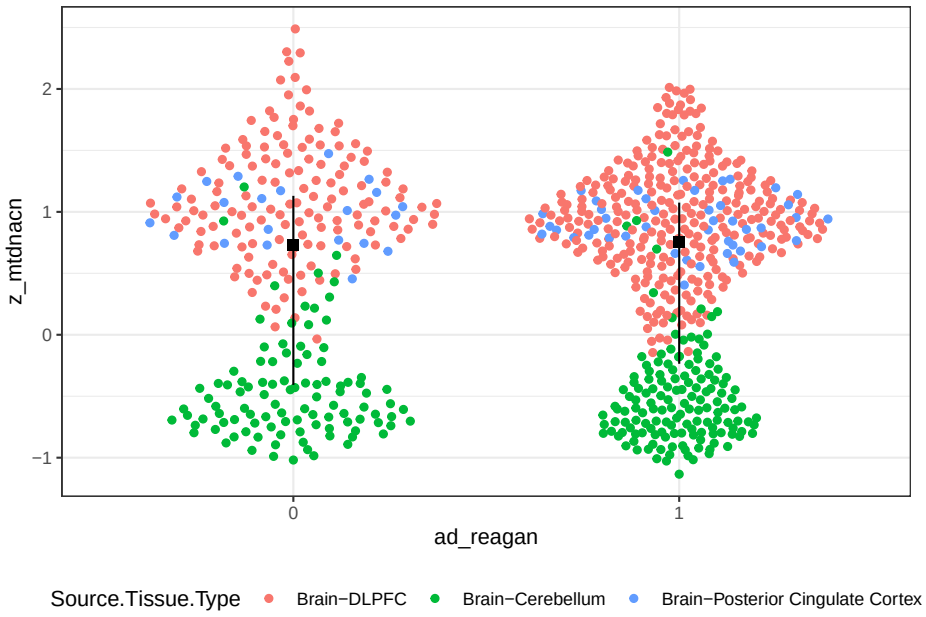
-0.076

0.304

-0.250

0.802

```
ggplot(rosmap_df, aes(x = ad_reagan, y = z_mtdnacn, colour = Source.Tissue.Type)) +
  geom_quasirandom() +
  geom_pointrange(mapping = aes(x = ad_reagan, y = z_mtdnacn),
                      show.legend = F, colour = 'black',
                     # size = 1,
                      position = position_dodge(width = 1),
                      shape = 15,
                      stat = "summary",
                      fun = median,
                      fun.min = function(z) {quantile(z,0.25)},
                      fun.max = function(z) {quantile(z,0.75)}) +
  theme_bw() + theme(legend.position = "bottom")
```



6.1.3 Clinical diagnosis

```
rosmap_clin_res <- glm(dx ~ z_mtdnacn + macro + age_death + msex + apoe4 + study + Source.Tissue.
```

Assocation of mtDNA with MMSE in ROSMAP

term

estimate

std.error

statistic

p.value

(Intercept)
-9.890
1.591
-6.215
5.1e-10 ***
z_mtdnacn
-0.453
0.248
-1.829
0.067 .
macroI
0.340
0.596
0.571
0.568
macroJ
0.092
0.396
0.233
0.816
macroK
0.991
0.437
2.266
0.023 *
macroT
-0.030
0.333
-0.091
0.928
macroU

0.340

0.283

1.202

0.23

macroV

0.502

0.483

1.039

0.299

macroW

0.595

0.733

0.812

0.417

macroX

-0.590

0.734

-0.804

0.421

age_death

0.110

0.017

6.479

9.3e-11 ***

msexM

-0.081

0.213

-0.379

0.704

apoe4e4+

1.429

0.243

5.881

4.1e-09 ***

studyROS

0.275

0.208

1.322

0.186

Source.Tissue.TypeBrain-Cerebellum

-0.426

0.417

-1.023

0.306

Source.Tissue.TypeBrain-Posterior Cingulate Cortex

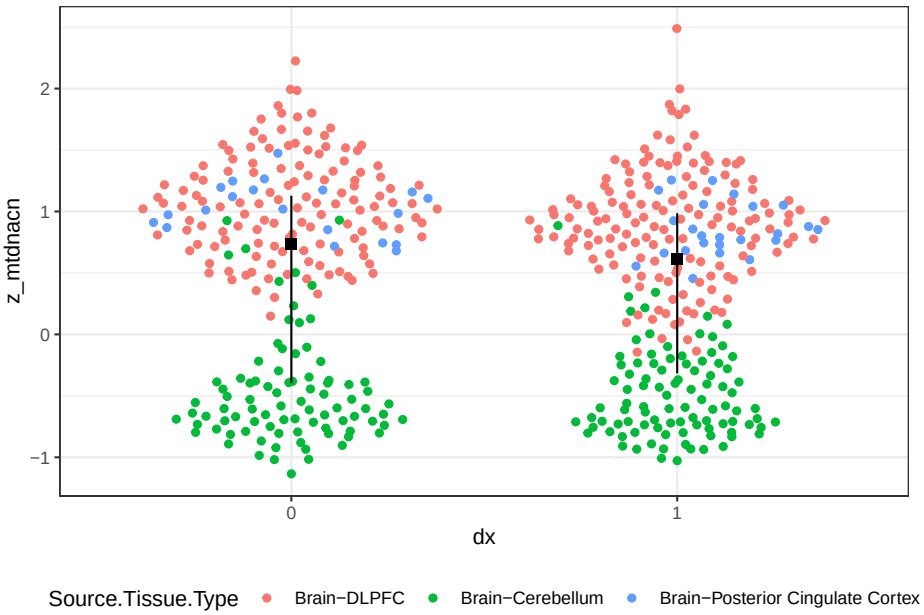
0.003

0.360

0.008

0.994

```
rosmap_df %>%
  filter(!is.na(dx)) %>%
  ggplot(., aes(x = dx, y = z_mtdnacn, colour = Source.Tissue.Type)) +
    geom_quasirandom() +
    geom_pointrange(mapping = aes(x = dx, y = z_mtdnacn),
                          show.legend = F, colour = 'black',
                         # size = 1,
                          position = position_dodge(width = 1),
                          shape = 15,
                          stat = "summary",
                          fun = median,
                          fun.min = function(z) {quantile(z,0.25)},
                          fun.max = function(z) {quantile(z,0.75)}) +
  theme_bw() + theme(legend.position = "bottom")
```



6.2 MSBB

6.2.1 Data Wrangling

- Exclude individules who are not non-hispanic whites
- Exclude individules with non-European haplogroups
- Dichotomize CERAD dx

```
msbb_df <- msbb  %>%
  filter(race == 'W') %>%
  filter(macro %in% c('H', 'V', 'J', 'T', 'U', 'K', 'I', 'W', 'X')) %>%
  mutate(cerad_dx = fct_recode(cerad, "0" = "Normal", "0" = "Possible", "1" = "Probable", "1" = '
```

6.2.2 NIA-Reagan

```
msbb_path_res <- glm(ad_reagan ~ z_mtdnacn + macro + aod + sex + apoe4,
                     family = "binomial", data = msbb_df)
```

Assocation of mtDNA with MMSE in msbb

term

estimate

std.error

statistic

p.value
(Intercept)
-2.204
1.542
-1.430
0.153
z_mtdnacn
-0.268
0.146
-1.837
0.066 .
macroI
15.877
1684.913
0.009
0.992
macroJ
0.088
0.498
0.177
0.86
macroK
0.342
0.372
0.920
0.358
macroT
0.866
0.630
1.374
0.169

macroU
0.767
0.561
1.367
0.172
macroV
0.936
0.615
1.522
0.128
macroW
16.010
1357.213
0.012
0.991
macroX
-0.201
1.040
-0.193
0.847
aod
0.026
0.017
1.517
0.129
sexM
0.411
0.341
1.206
0.228
apoe4e4+

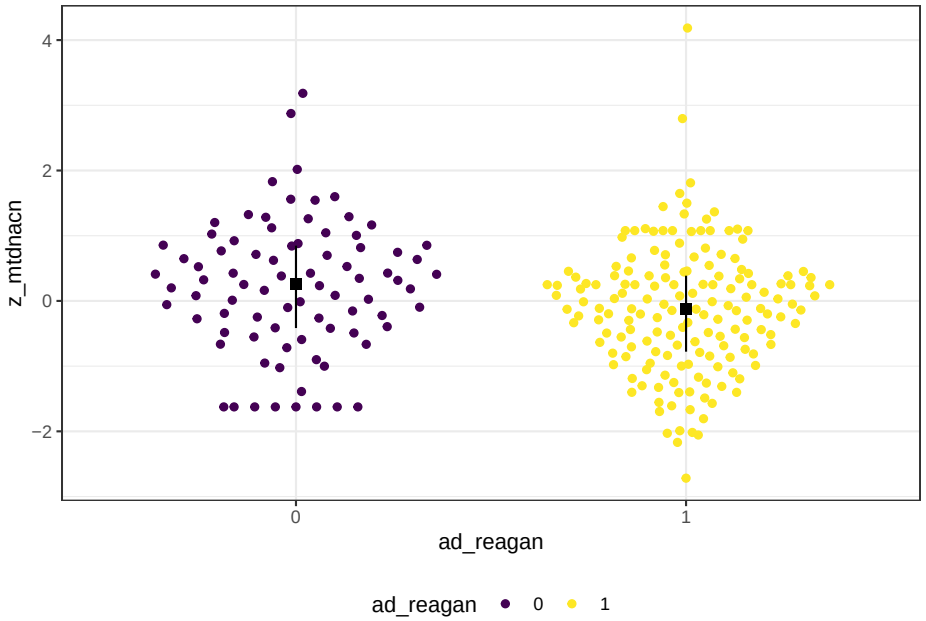
0.760

0.320

2.375

0.018 *

```
msbb_df %>%
  filter(!is.na(ad_reagan)) %>%
    ggplot(., aes(x = ad_reagan, y = z_mtdnacn, colour = ad_reagan)) +
      geom_quasirandom() +
      geom_pointrange(mapping = aes(x = ad_reagan, y = z_mtdnacn),
                      show.legend = F, colour = 'black',
                     # size = 1,
                      position = position_dodge(width = 1),
                      shape = 15,
                      stat = "summary",
                      fun = median,
                      fun.min = function(z) {quantile(z,0.25)},
                      fun.max = function(z) {quantile(z,0.75)}) +
      theme_bw() + theme(legend.position = "bottom")
```



6.2.3 CERAD

```
msbb_cerad_res <- glm(cerad_dx ~ z_mtdnacn + macro + aod + sex + apoe4,
                      family = "binomial", data = msbb_df)
```

Assocation of mtDNA with MMSE in msbb

term

estimate

std.error

statistic

p.value

(Intercept)

-1.680

1.500

-1.120

0.263

z_mtdnacn

-0.629

0.151

-4.155

3.3e-05 ***

macroI

15.049

840.116

0.018

0.986

macroJ

-0.182

0.419

-0.434

0.664

macroK

0.748

0.359

2.081

0.037 *

macroT
0.934
0.597
1.564
0.118
macroU
2.208
0.782
2.825
0.005 **
macroV
1.002
0.576
1.740
0.082 .
macroW
0.854
1.261
0.677
0.498
macroX
0.252
1.026
0.246
0.806
aod
0.017
0.017
1.028
0.304
sexM

0.338

0.312

1.082

0.279

apoe4e4+

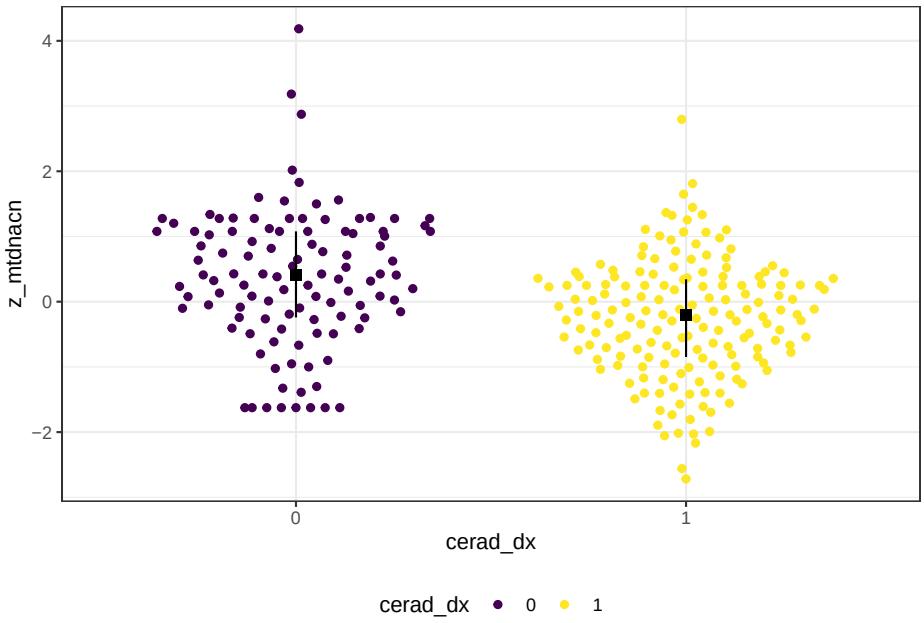
0.358

0.291

1.233

0.217

```
msbb_df %>%
  filter(!is.na(cerad_dx)) %>%
    ggplot(., aes(x = cerad_dx, y = z_mtdnacn, colour = cerad_dx)) +
      geom_quasirandom() +
      geom_pointrange(mapping = aes(x = cerad_dx, y = z_mtdnacn),
                      show.legend = F, colour = 'black',
                     # size = 1,
                      position = position_dodge(width = 1),
                      shape = 15,
                      stat = "summary",
                      fun = median,
                      fun.min = function(z) {quantile(z,0.25)},
                      fun.max = function(z) {quantile(z,0.75)}) +
      theme_bw() + theme(legend.position = "bottom")
```


4
2
0
−2
z_mtdnacn
0
1
cerad_dx
cerad_dx
0
1

Chapter 7

AAIC Abstract

Mitochondrial DNA copy number is associated with cognitive impairment

Background: Increasing evidence has implicated mitochondrial dysfunction in the pathogenesis of Alzheimer's Disease. Mitochondria contain their own DNA outside of the nuclear genome, with every cell having between 100-10,000 copies of mtDNA. Mitochondrial DNA copy number (mtDNA-CN) has been used as a surrogate measure of mitochondrial function, with reduced mtDNA-CN associated with age-related diseases. The aim of this study was to evaluate the association of mtDNA-CN with cognitive impairment.

```
rosmap.raw  <- readRDS('output/rosmap.RData')
rosmap <- rosmap.raw %>%
  filter(Source.Tissue.Type %in% c("Brain-DLPFC", "Brain-Posterior Cingulate Cortex", "Brain-Cere
  filter(organ == 'brain') %>%
  filter(macro %in% c('H', 'V', 'J', 'T', 'U', 'K', 'I', 'W', 'X')) %>%
  filter(!is.na(macro)) %>%
  filter(!is.na(cts_mmse30_lv)) %>%
  filter(!is.na(apoe4)) %>%
  mutate(Source.Tissue.Type = str_replace(Source.Tissue.Type, "Brain-", ""),
         Source.Tissue.Type = str_replace(Source.Tissue.Type, "Posterior Cingulate Cortex", "PCC"
  select(cts_mmse30_lv, niareagansc, dcfdx_lv, cts_mmse30_lv, pmi, study, age_death, msex,
         Source.Tissue.Type, organ, apoe4, z_mtdnacn, mtcn_avg, macro)
```

```
msbb.raw <- readRDS('output/msbb.rds')

msbb <- msbb.raw  %>%
  filter(race == 'W') %>%
  filter(macro %in% c('H', 'V', 'J', 'T', 'U', 'K', 'I', 'W', 'X'))
```

Methods: We evaluated the association of mtDNA-CN with the extended Clin-

ical Dementia Rating (CDR) scale in the Mount Sinai Brain Bank (MSBB) and with mini mental state exam (MMSE) in the Religious Orders Studies and the Memory Aging Project (ROSMAP). Relative mtDNA-CN was estimated as the ratio of mitochondrial genomes to nuclear genomes in 1025 non-Hispanic white subjects (MSBB = 277; ROSMAP = 748) using whole-genome sequencing data generated from DNA isolated from post-mortem brain tissue (MSBB: prefrontal cortex; ROSMAP: dorsolateral prefrontal cortex [DLPFC], posterior cingulate cortex [PCC], or cerebellum). Linear regression adjusting for age of death, sex, APOE, study, mitochondrial haplogroup and source tissue were used to evaluate the association of mtDNA-CN with cognitive impairment.

```
## CDR analysis
cdr_res <- lm(cdr ~ z_mtdnacn  + aod + sex + apoe4 + macro, data = msbb)
cdr_tab <- tidy(cdr_res)

## MMSE analysis
mmse_res <- lm(cts_mmse30_lv ~ z_mtdnacn  + age_death + msex + apoe4 + study + Source.
mmse_tab <- tidy(mmse_res)
```

Results: In the MSBB, a one standard deviation decrease (1 s.d. 343) in mtDNA-CN was associated with a higher CDR score (β(se) = 0.7 (0.1), p = 3.41e-12, **??**). Similarly, in ROSMAP a one standard deviation decrease (1 s.d. 1063) in mtDNA-CN was associated with a lower MMSE score (β(se) = -4.02 (0.75), p = 1.07e-07, **??**).

Assocation of mtDNA with CDR in MSBB

term

estimate

std.error

statistic

p.value

(Intercept)

1.859

1.037

1.793

0.074

z_mtdnacn

-0.705

0.097

-7.301

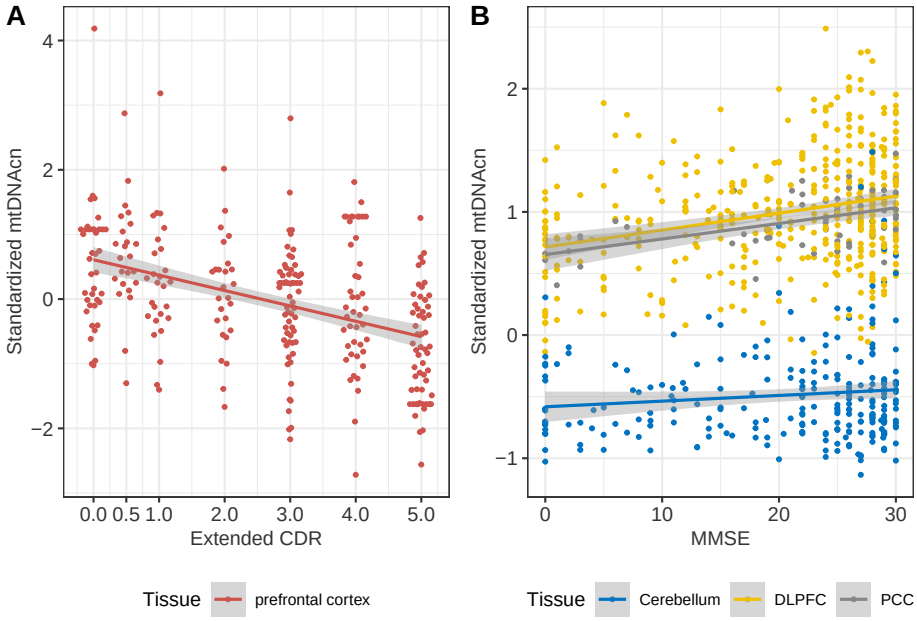


Figure 7.1: Relationship between mtDNAcn and CDR & MMSE in MSBB and ROSMAP

3.41e-12

aod

0.002

0.012

0.192

0.848

sexM

0.084

0.215

0.388

0.698

apoe4e4+

0.724

0.200

3.621

3.52e-04

Haplogroup

macroI

0.938

0.918

1.022

0.308

macroJ

0.866

0.308

2.809

0.005

macroK

0.087

0.255

0.343

0.732

macroT

0.818

0.398

2.055

0.041

macroU

0.470

0.361

1.302

0.194

macroV

0.748

0.395

1.896

0.059

macroW

0.801

0.916

0.874

0.383

macroX

-0.769

0.794

-0.968

0.334

Assocation of mtDNA with MMSE in ROSMAP

term

estimate

std.error

statistic

p.value

(Intercept)

49.199

4.450

11.056

2.19e-26

z_mtdnacn

4.017

0.748

5.368

1.07e-07

age_death

-0.281

0.049

-5.744

1.36e-08

msexM

0.823

0.674

1.222

0.222

apoe4e4+

-4.529

0.727

-6.230

7.89e-10

studyROS

-1.110

0.644

-1.722

0.085

Tissue

Source.Tissue.TypeDLPFC

-5.443

1.310

-4.156

3.62e-05

Source.Tissue.TypePCC

-3.561

1.604

-2.220

0.027

Haplogroup

macroI

1.062

1.737

0.611

0.541

macroJ

-0.959

1.181

-0.812

0.417

macroK

-1.863

1.241

-1.502

0.134

macroT

-0.033

1.089

-0.031

0.976

macroU

-1.209

0.919

-1.316

0.188

macroV

-1.873

1.479

-1.267

0.206

macroW

-5.087

2.514

-2.023

0.043

macroX

2.000

2.629

0.761

0.447

Conclusion: Mitochondrial dysfunction as measured by mtDNA-CN is associated with worse cognitive performance, suggesting that mitochondrial function plays a role in the pathogenesis of Alzheimer's Disease. However, further research is needed to determine if mitochondrial dysfunction causes, mediates, or is a by-product of AD pathogenesis, in particular whether neuronal loss is an unobserved confounder that could be driving the observed associations.

Chapter 8

Final Words

We have finished a nice book.

Bibliography

Jager, P. L. D., Ma, Y., McCabe, C., Xu, J., Vardarajan, B. N., Felsky, D., Klein, H.-U., White, C. C., Peters, M. A., Lodgson, B., Nejad, P., Tang, A., Mangravite, L. M., Yu, L., Gaiteri, C., Mostafavi, S., Schneider, J. A., and Bennett, D. A. (2018). A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Scientific Data*, 5(1):180142.

Mirra, S. S., Heyman, A., McKeel, D., Sumi, S. M., Crain, B. J., Brownlee, L. M., Vogel, F. S., Hughes, J. P., Belle, G. v., and Berg, L. (1991). The consortium to establish a registry for alzheimer's disease (cerad). *Neurology*, 41(4):479–479.

Wang, M., Beckmann, N. D., Roussos, P., Wang, E., Zhou, X., Wang, Q., Ming, C., Neff, R., Ma, W., Fullard, J. F., Hauberg, M. E., Bendl, J., Peters, M. A., Logsdon, B., Wang, P., Mahajan, M., Mangravite, L. M., Dammer, E. B., Duong, D. M., Lah, J. J., Seyfried, N. T., Levey, A. I., Buxbaum, J. D., Ehrlich, M., Gandy, S., Katsel, P., Haroutunian, V., Schadt, E., and Zhang, B. (2018). The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Scientific Data*, 5(1):180185.