HS-BMMB-009

Neuromelanin and Parkinson's disease: Systematic analysis of age-dependent and population-specific gene expression signatures

Q1: Problem or Question

- Parkinson's Disease (PD) is caused by death of the dopaminergic (DA) neurons of the neuromelanin-containing **substantia nigra pars compacta** (**SNpc**)
  - DA neurons lacking neuromelanin (NM), such as those in the **ventral tegmental area** (**VTA**), do not die in PD
  - Studies have shown that aging is closely linked to NM and PD

Research Question: By which genes and biological processes does neuromelanin contribute to the pathogenesis of Parkinson's disease in an age-dependent manner?

Q2: Framework

Goal: To identify genes and biological processes involved in age-dependent neuromelanin-linked neurotoxicity that contribute to pathogenesis of Parkinson's disease

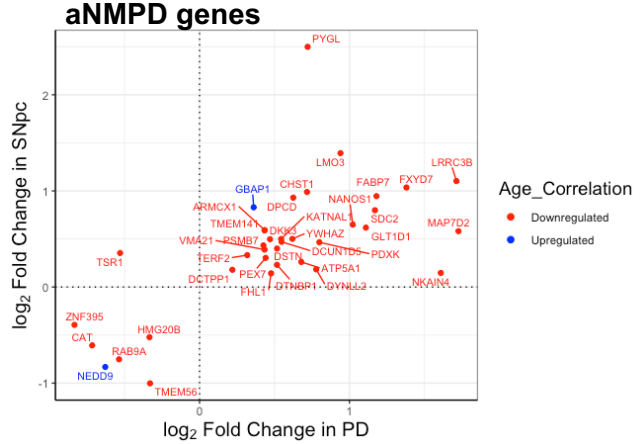
Identifying sets of genes whose expression differ…

a) With age (expression-age correlations) in the substantia nigra
b) In presence of neuromelanin (SNpc vs VTA, differential expression analysis with correction for cell type)
c) In Parkinson's (differential expression analysis)

… and finding genes that fit in all three categories (**aNMPD genes**)

Q3: Findings

1. Genes differentially expressed in microglia of SNpc exhibited enrichment for **aging** and **Parkinson's disease** ($p < 0.05$)
2. RAB9A, YWHAZ, DTNBP1, CAT (aNMPD genes) interact with PD-implicated proteins (mean interaction score > 0.4)



Q4: Conclusions

1. Microglial cells are the link between aging, NM, and Parkinson's disease
   - Microglia-specific expression changes in SNpc contribute to aging and PD pathogenesis in presence of NM

2. Proteins RAB9A, YWHAZ, DTNBP1, CAT could serve as preventative/therapeutic drug targets
   - Prevent/treat NM-linked neurotoxicity
   - Use to construct more accurate animal models of PD

Introduction

Parkinson's Disease

- Incidence of Parkinson's disease (PD) expected to increase by 30% by 2030 (Kouli, Torsney, & Kuan, 2018)
- Caused by death of dopaminergic (DA) neurons of the **substantia nigra pars compacta (SNpc**) linked to the aggregation of **α-synuclein** (Vila, 2019)
- **Age is the greatest risk factor** for PD
- **Only melanized DA neurons significantly lost** in PD [2,3] (Vila, 2019; Haining & Achat-Mendes, 2017)
- DA neurons in SNpc contain high levels of neuromelanin (are melanized)
- DA neurons in **ventral tegmental area (VTA**), lack neuromelanin, not lost in PD (Vila, 2019)

Unknowns about Parkinson's

- **Molecular mechanism** linking **PD** with **age** and **Neuromelanin** (**NM**)
- Initial trigger of neurotoxic mechanisms/neuron loss in PD (could be related to NM)
- **Drug targets for disease-modifying therapies** involving NM

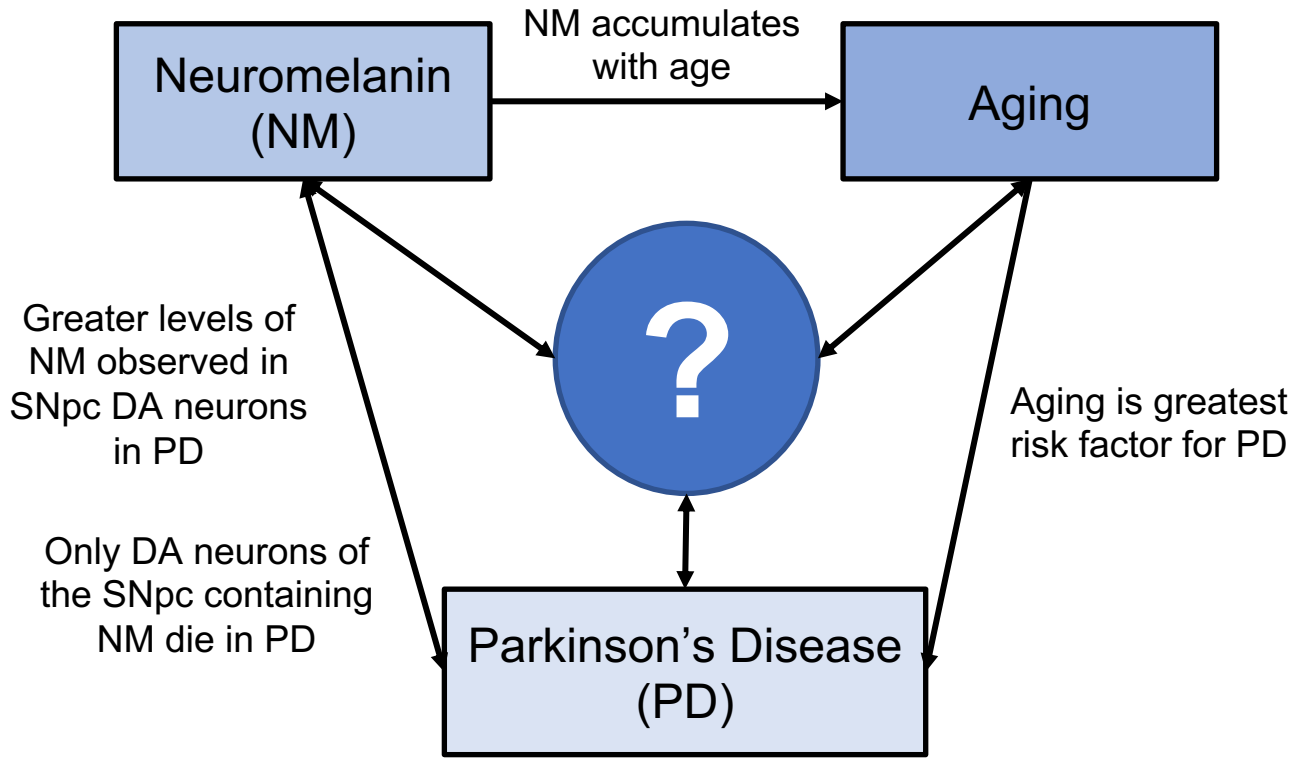


Figure 1. Graphical illustration of observations in current literature that suggest an as-of-yet unknown link between neuromelanin (NM), aging, and Parkinson's disease (PD) exists.

Research Question: By which genes and biological processes does neuromelanin contribute to the pathogenesis of Parkinson's disease in an age-dependent manner?

Unless otherwise specified, all figures are produced by the student researcher

Framework

Goal: To identify genes and biological processes involved in age-dependent neuromelanin-linked neurotoxicity that may contribute to pathogenesis of Parkinson's disease

1. **Expression-Age Correlations** → address age factor
   - Identify genes significantly correlated with age ($\alpha = 0.05$)
2. **Melanized Dopaminergic Brain Region analysis** (**mDBR**) → address NM factor
   - Identify genes differentially expressed in presence of NM ($\alpha = 0.05$)
   - Differential expression analysis between melanized SNpc and non-melanized VTA
   - Observe cell-type specific expression changes using population-specific expression analysis (PSEA)
3. **Overlap and Validation in PD**
   - Overlap: Identify genes that are both correlated with age and differentially expressed in SNpc
   - Confirm overlapping genes also differentially expressed in SN of PD brains ($\alpha = 0.05$)

Relevant Concepts

Differential Expression Analysis:
- Identify genes exhibiting expression changes in SNpc relative to VTA (**mDBR**) and in PD relative to controls (**Validation in PD**)
- Two-sided, Two-tailed T-tests
- Benjamini-Hochberg (BH) procedure to correct p-values → minimize false discovery rate (FDR)

Population-Specific Expression Analysis (PSEA):
- Correct for cell-type population size differences between SNpc and VTA (**mDBR**) in differential expression analysis
- Dopaminergic neurons, GABAergic neurons, Glutaminergic neurons, Oligodendrocytes, Microglia, Astrocytes
- Regress gene expression on expression of reference genes, using interaction regressor for brain region

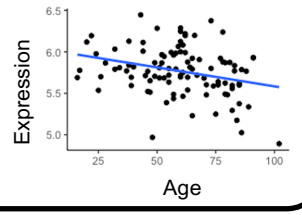
Gene Ontology Term Enrichment and Protein-Protein Interaction Networks:
- Identify biological processes, molecular functions, and cellular components associated with gene sets
- Characterize interactions of proteins within gene sets with each other and with PD-implicated genes previously noted in the literature

Framework (continued)

1. Expression-Age Correlations

Gene Expression in Substantia Nigra of Healthy Individuals (UK Brain Expression Consortium)

Expression-Age Pearson's Correlations in R



Age-Correlated Genes (ACGs)

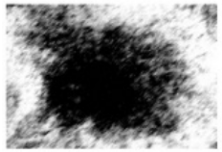
Validation in PD patients
Differential Expression Analysis in SN of PD patients vs controls (3 datasets, analyzed with GEO2R)

Overlap: **aNMPD genes**
Genes involved in **age-dependent NM-linked neurotoxicity in PD**

2. Melanized Dopaminergic Brain Region Analysis

Gene Expression Data (Allen Human Brain Atlas)

Substantia nigra pars compacta (SNpc)
Contains neuromelanin, dopaminergic



Ventral tegmental area (VTA)
No neuromelanin, dopaminergic

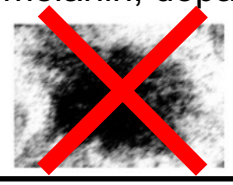
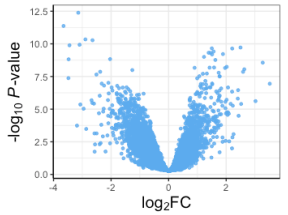
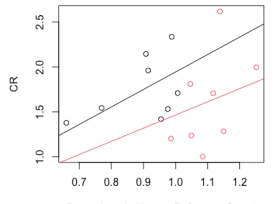


Image from Munoz et al., 2012



Differential Expression Analysis

Population-specific expression analysis
Multiple Linear Regression in R



mDBR genes
Genes differentially expressed in **m**elanized **d**opaminergic **b**rain **r**egions (SNpc relative to VTA)

Protein-Protein Interaction Networks

Separate by cell type
Dopaminergic, GABAergic, and Glutamatergic neurons, Astrocytes, Microglia, Oligodendrocytes

Gene Ontology (GO) Term Enrichment Analysis

Findings

1. Population-specific expression changes in SNpc (Microglia)

Microglia-specific genes exhibited enrichment for aging and Parkinson's disease

PSEA: Applying linear models to gene expression data to identify genes whose expression is different in certain cell types of SNpc compared to the same cell type in the VTA

Population-specific expression changes in microglia of SNpc relative to VTA (Fig. 2)

- GO term enrichment of genes exhibiting microglia-specific expression changes in SNpc:
  - **Aging** (Fold enrichment = 1.51, BH-corrected p = 0.0237)
  - **Parkinson's disease** (Fold enrichment = 2.36, BH-corrected p = 0.0188)
  - **Response to iron(II) ion** (Fold enrichment = 4.17, BH-corrected p = 0.0128)
  - Included PD-implicated genes **SNCA** (codes for α-synuclein), **PINK1**, HTRA2, CHCHD2

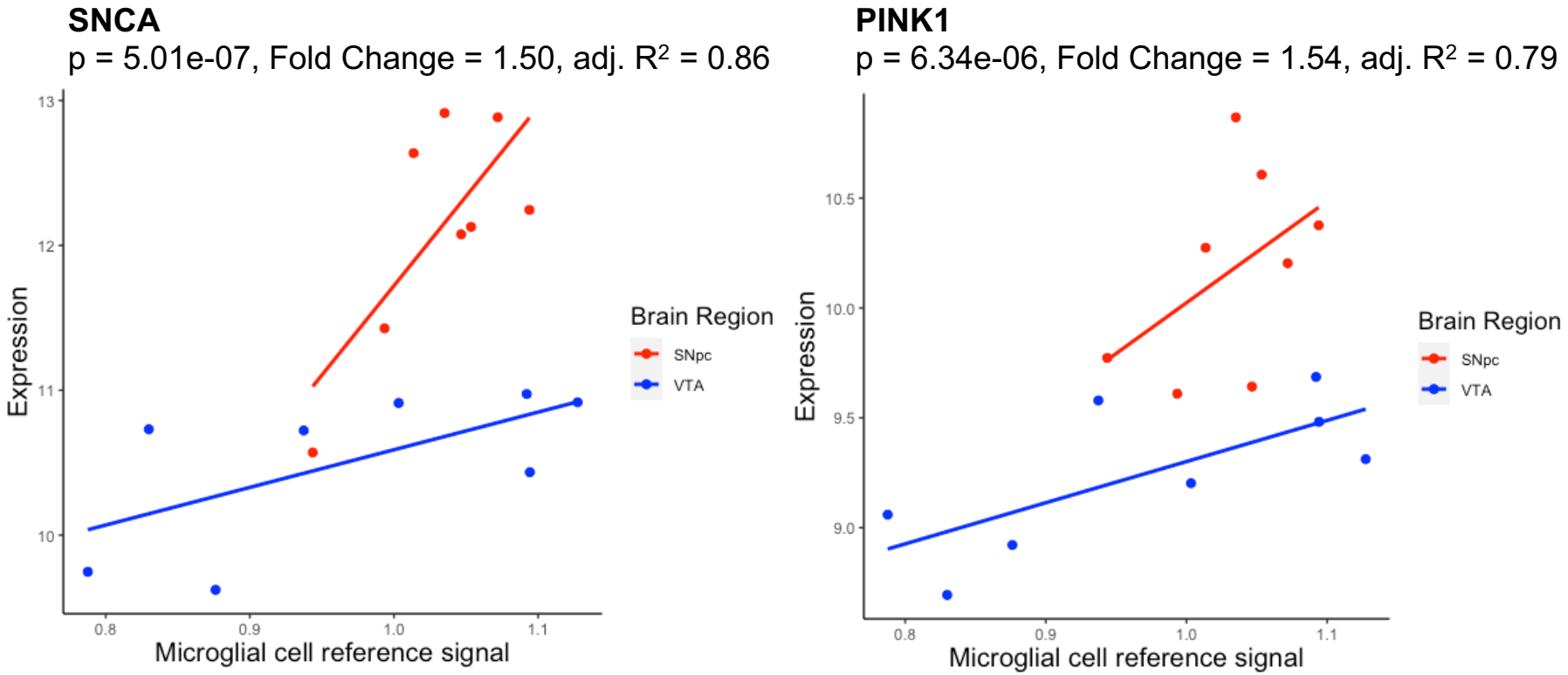


Figure 2. Scatterplots of microglial cell-specific expression of SNCA and PINK1 in the SNpc and VTA. Red indicates expression in the SNpc, and blue indicates expression in the VTA. The greater slopes of the SNpc lines relative to the VTA lines indicate that SNCA and PINK1 are upregulated in the microglia of the SNpc. The high adjusted R^2 values indicate that the microglia-specific expression accounts for the majority of the variation in expression of these genes. Data from Allen Human Brain Atlas, graphs produced by student researcher.

Findings (continued)

2. Genes involved in age-dependent neuromelanin-linked neurotoxicity in PD (aNMPD genes)

37 genes identified that may contribute to age-dependent NM-linked neurotoxicity (Fig. 3)

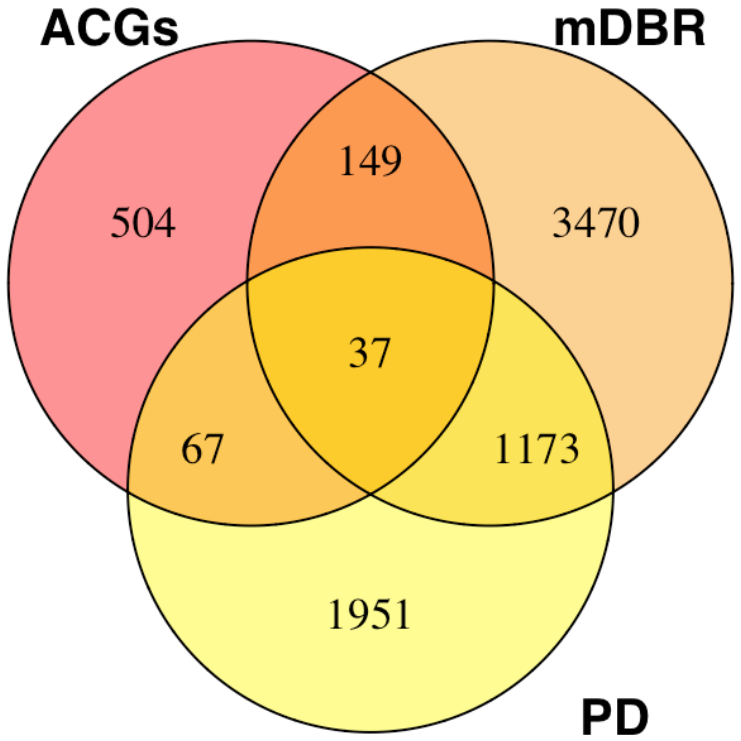


Figure 3. Venn diagram of genes that were found to be age-correlated in the SN (ACGs), differentially expressed in the SNpc (mDBR), and differentially expressed in PD. The 37 genes in the center are aNMPD genes that may contribute to age-dependent neuromelanin-linked neurotoxicity in PD.

Patterns of Differential Expression and Correlation with Age (Fig. 4)

- Most aNMPD genes exhibited similar patterns of differential expression in SNpc and PD
- Few aNMPD genes exhibited similar patterns of differential expression and correlation with age
- aNMPD genes enriched for identical protein binding molecular function (fold enrichment = 1.44, p = 0.0399)

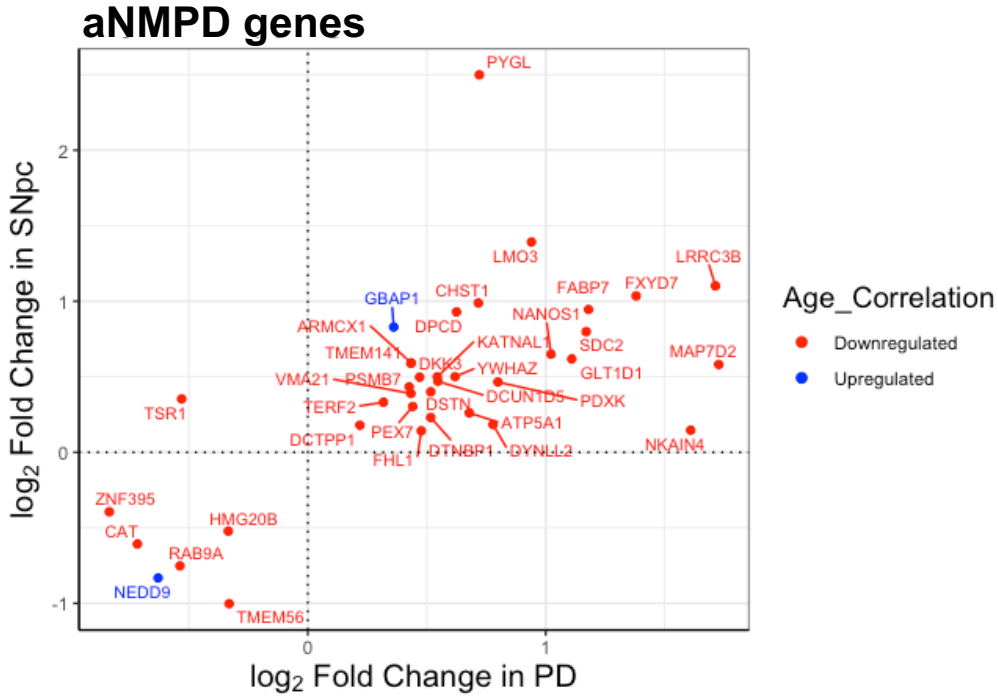


Figure 4. Plot comparing differential expression of 37 aNMPD genes in the SNpc and in PD and their correlations with age in the SN. Red indicates genes that were downregulated with age; blue indicates genes that were upregulated with age. Genes in first quadrant were upregulated in both the SNpc and the PD, and genes in the third quadrant were downregulated in both the SNpc and PD. Graph produced by student researcher.

Findings (continued)

2. Genes involved in age-dependent neuromelanin-linked neurotoxicity in PD (aNMPD genes)

RAB9A, YWHAZ, DTNBP1, CAT exhibited significant interactions with PD-implicated genes

- aNMPD genes: RAB9A, YWHAZ, DTNBP1, CAT (Table 1)
  - Interactions with PD-implicated proteins (mean interaction score > 0.4, STRING DB)
  - Associated with NM and PD-related GO terms (melanosome, reactive oxygen species metabolic process)
- RAB9A is required for trafficking of melanogenic enzymes to the melanosome
- CAT protects against hydrogen peroxide, also promotes T cell growth

Table 1. Statistics from analyses of RAB9A, YWHAZ, DTNBP1, CAT. All p-values listed are BH-corrected. Full table of select aNMPD genes is available in Supplementary Table 1 and all aNMPD genes are listed in Supplementary Table 2. See Supplementary Figure 1 for full protein network of aNMPD genes and PD-implicated genes.

Gene	Correlation with Age (*r*)	P-value of Correlation with Age	Fold change in SNpc	P-value of fold change in SNpc	Fold change in PD	P-value of fold change in PD	Mean Interaction Score*	Gene Ontology (GO) terms
RAB9A	–0.39	0.0087	0.59	0.0018	0.69	0.028	0.493	Melanosome
YWHAZ	–0.32	0.042	1.4	0.0080	1.54	0.0038	0.717	Melanosome
DTNBP1	–0.33	0.032	1.17	0.0039	1.43	0.018	0.902	Melanosome membrane
CAT	–0.31	0.044	0.66	0.0064	0.61	0.038	0.581	Aging, Reactive oxygen species metabolic process

*Mean Interaction Score is the average interaction score with proteins previously implicated in PD in the literature as provided by STRING DB. These proteins are listed in Supplementary Table 3.

Conclusions

1. Conclusions from population-specific expression changes in SNpc

Genes differentially expressed in microglia of NM-containing SNpc exhibited enrichment for **aging** and **Parkinson's disease** and included **SNCA**

Suggests **microglial cells in SNpc contribute to aging and PD pathogenesis** (by α-synuclein) **in presence of NM**

Xuan et al. 2011 found increased expression of α-synuclein has been associated with NM accumulation in humans

Microglia contribute to this increased expression of α-synuclein in presence of NM

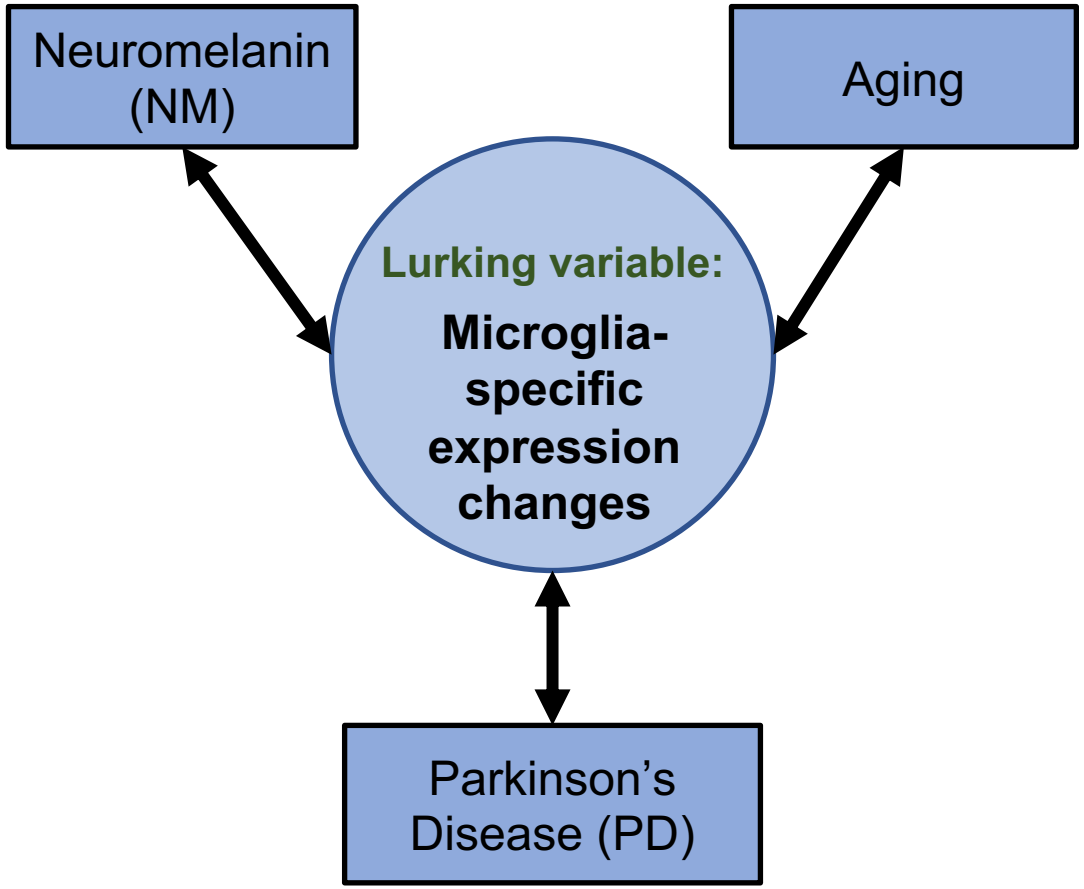


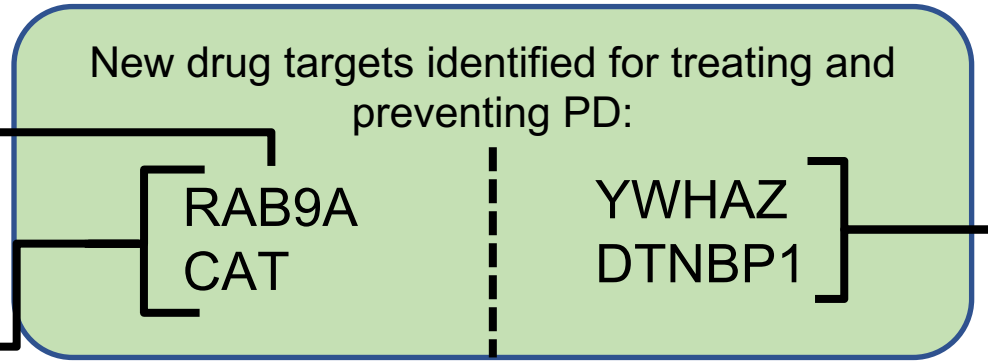
Figure 5. Microglial cell-specific gene expression in the presence of neuromelanin (in the SNpc) contributes to both aging and the pathogenesis of Parkinson's disease. This was concluded from the observation that genes differentially expressed in microglia of the SNpc were enriched for the GO terms aging and Parkinson's disease, and also included genes associated with the melanosome GO term, namely YWHAZ and DTNBP1 (Supplementary Table 2).

Conclusions (continued)

2. Conclusions from identification of aNMPD genes

The observed downregulation of RAB9A may be a protective mechanism, as overexpression of RAB9 has been linked to increased α-synuclein presence (Dinter et al., 2016)

Preventative drug targets: age-dependence aligns with differential expression



Therapeutic targets for remediation of NM-linked neurotoxicity

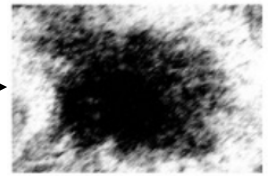
Applications and Future Research

- Currently a rat model of PD based on NM neurotoxicity does exist (Carballo-Carbajal et al., 2019), although it is inaccurate because it is based on the injection of human tyrosinase into rat SNs to induce NM formation
- **Using the genes identified in study to construct animal models of PD based on natural NM-linked neurotoxicity will increase model accuracy (Fig. 5)**

Image from NPR.org

Knockout of RAB9A, CAT
Overexpress YWHAZ and DTNBP1

NM formation?



PD-like symptoms and pathology?

Figure 6. Theoretical construction of a future study that would evaluate the efficacy of using select aNMPD genes to generate to more accurate rodent model of PD. This model would be more pathologically accurate because it is based off neuromelanin-linked neurotoxicity rather than using external neurotoxins or genes to induce DA neuron death Such a study would involve altering the expression of RAB9A, CAT, YWHAZ, and DTNBP1 in accordance with their expression differences observed with age, in the SNpc, and in PD patients in this study.

References

1. Carballo-Carbajal, I., Laguna, A., Romero-Giménez, J., Cuadros, T., Bové, J., Martinez-Vicente, M., Parent, A., Gonzalez-Sepulveda, M., Peñuelas, N., Torra, A., Rodríguez-Galván, B., Ballabio, A., Hasegawa, T., Bortolozzi, A., Gelpi, E., & Vila, M. (2019). Brain tyrosinase overexpression implicates age-dependent neuromelanin production in Parkinson's disease pathogenesis. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-08858-y

2. Haining, R. L., & Achat-Mendes, C. (2017). Neuromelanin, one of the most overlooked molecules in modern medicine, is not a spectator. *Neural Regeneration Research*, *12*(3), 372–375. https://doi.org/10.4103/1673-5374.202928

3. Kouli, A., Torsney, K. M., & Kuan, W. (2018). Parkinson's Disease: Etiology, Neuropathology, and Pathogenesis. 3–26. https://doi.org/https://doi.org/10.15586/codonpublications.parkinsonsdisease.2018.ch1

4. Keo, A., Mahfouz, A., Ingrassia, A. M. T., Meneboo, J. P., Villenet, C., Mutez, E., Comptdaer, T., Lelieveldt, B. P. F., Figeac, M., Chartier-Harlin, M. C., van de Berg, W. D. J., van Hilten, J. J., & Reinders, M. J. T. (2020). Transcriptomic signatures of brain regional vulnerability to Parkinson's disease. *Communications Biology*, *3*(1), 1–12. https://doi.org/10.1038/s42003-020-0804-9

5. Vila, M. (2019). Neuromelanin, aging, and neuronal vulnerability in Parkinson's disease. *Movement Disorders*, *34*(10), 1440–1451. https://doi.org/10.1002/mds.27776

6. Xuan, Q., Xu, S. L., Lu, D. H., Yu, S., Zhou, M., Uéda, K., Cui, Y. Q., Zhang, B. Y., & Chan, P. (2011). Increase expression of a-synuclein in aged human brain associated with neuromelanin accumulation. *Journal of Neural Transmission*, *118*(11), 1575–1583. https://doi.org/10.1007/s00702-011-0636-3

7. Zhang, W., Phillips, K., Wielgus, A. R., Liu, J., Albertini, A., Zucca, F. A., Faust, R., Qian, S. Y., Miller, D. S., Chignell, C. F., Wilson, B., Jackson-Lewis, V., Przedborski, S., Joset, D., Loike, J., Hong, J. S., Sulzer, D., & Zecca, L. (2011). Neuromelanin activates microglia and induces degeneration of dopaminergic neurons: Implications for progression of parkinson's disease. *Neurotoxicity Research*. https://doi.org/10.1007/s12640-009-9140-z

8. Dinter, E., Saridaki, T., Nippold, M., Plum, S., Diederichs, L., Komnig, D., Fensky, L., May, C., Marcus, K., Voigt, A., Schulz, J. B., & Falkenburger, B. H. (2016). Rab7 induces clearance of α-synuclein aggregates. *Journal of Neurochemistry*. https://doi.org/10.1111/jnc.13712

Virtual Logbook Excerpts 1

```
1168 compute_all_age_correlations <- function(exp, age_vector) {
1169
1170   # Data frame to store correlation data
1171   all_cor <- matrix(nrow = dim(exp)[1], ncol = 4) %>% data.frame
1172   colnames(all_cor) <- c("Gene", "Correlation Coefficient", "Raw p-value", "BH-corrected p-value")
1173   rownames(all_cor) <- rownames(exp)
1174   all_cor[,"Gene"] <- rownames(exp)
1175
1176   # Loop that performs Pearson's correlations
1177   for (i in 1:dim(exp)[1]) {
1178     cor <- cor.test(as.numeric(exp[i, ]),
1179                     age_vector,
1180                     method = "pearson")
1181     all_cor[i, "Correlation Coefficient"] <- cor[["estimate"]][["cor"]]
1182     all_cor[i, "Raw p-value"] <- cor[["p.value"]]
1183   }
1184
1185   # Correcting p-values using Benjamini-Hochberg Procedure
1186   all_cor[, "BH-corrected p-value"] <- p.adjust(all_cor[, "Raw p-value"], method = "BH")
1187   return(all_cor)
1188
1189 }
```

Figure 7. Logbook excerpt of function used to compute expression-age correlations.

Virtual Logbook Excerpts 2

```r
library(PSEA)
library(tidyverse)

groups_mDBR <- c(1, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0)

VTA_samples <- AHBA_expr_SNpc_VTA[, c(3, 4, 7, 8, 10, 12, 14, 16)]

# Astrocyte
astro_probesets_mDBR <- list("GFAP", "AQP4", "GJA1")
astro_reference_mDBR <- marker(as.matrix(AHBA_expr_SNpc_VTA[, 1:16]), astro_probesets_mDBR)
astro_difference_mDBR <- groups_mDBR * astro_reference_mDBR

# Oligodendrocyte
oligo_probesets_mDBR <- list("MBP", "MOBP", "MOG", "MAG")
oligo_reference_mDBR <- marker(as.matrix(AHBA_expr_SNpc_VTA[, 1:16]), oligo_probesets_mDBR)
oligo_difference_mDBR <- groups_mDBR * oligo_reference_mDBR

# Microglia
micro_probesets_mDBR <- list("AIF1", "ITGAM")
micro_reference_mDBR <- marker(as.matrix(AHBA_expr_SNpc_VTA[, 1:16]), micro_probesets_mDBR)
micro_difference_mDBR <- groups_mDBR * micro_reference_mDBR

# Dopaminergic neuron
DA_probesets_mDBR <- list("TH", "SLC6A3")
DA_reference_mDBR <- marker(as.matrix(AHBA_expr_SNpc_VTA[, 1:16]), DA_probesets_mDBR)
DA_difference_mDBR <- groups_mDBR * DA_reference_mDBR

# Glutamatergic neuron
Glut_probesets_mDBR <- list("SLC17A7", "SLC17A6")
Glut_reference_mDBR <- marker(as.matrix(AHBA_expr_SNpc_VTA[, 1:16]), Glut_probesets_mDBR)
Glut_difference_mDBR <- groups_mDBR * Glut_reference_mDBR

# GABAergic neuron
GABA_probesets_mDBR <- list("GAD1", "GAD2", "SLC32A1")
GABA_reference_mDBR <- marker(as.matrix(AHBA_expr_SNpc_VTA[, 1:16]), GABA_probesets_mDBR)
GABA_difference_mDBR <- groups_mDBR * GABA_reference_mDBR
```

```r
PSEA_cell_type_specific_mDBR <- function(reference, difference) {

  AHBA_expr_SNpc_VTA_tranposed <- t(AHBA_expr_SNpc_VTA[, 1:16])

  # Construct all models
  models <- list()
  for (i in rownames(AHBA_expr_SNpc_VTA[, 1:16])) {
    models[[i]] <- lm(AHBA_expr_SNpc_VTA_tranposed[,i] ~ reference + difference)
  }

  # Extract model info (p-values, coefficients, adjusted R2)
  reg <- c("(Intercept)", "reference", "difference")
  coef <- coefmat(models, reg)
  p <- pvalmat(models, reg)
  summaries <- lapply(models, summary)
  adj_R2 <- slt(summaries, 'adj.r.squared')

  # Filter for models with R2 > 0.50, intercept is less than 75% of average expression
  avg_expr <- apply(VTA_samples, 1, mean)
  filt <- adj_R2 > 0.5 & (coef[,1] / avg_expr) < 0.75

  # Select models with p for both regressors < 0.5 and reference signal regressor > 0
  specific <- which(filt & p[, 2] < 0.05 & p[, 3] < 0.05 & coef[, 2] > 0)
  specific <- data.frame(specific)

  specific$Gene <- rownames(specific)

  coef <- data.frame(coef)
  coef$num <- as.numeric(rownames(coef))

  p <- data.frame(p)
  p$num <- as.numeric(rownames(p))

  # Put all selected models and p-values and fold changes into one data frame
  specific <- specific %>% inner_join(p, by = c("specific" = "num"))
  specific <- specific %>% inner_join(coef, by = c("specific" = "num"))
  specific <- specific[,-c(1,3,4,6)]
  specific$fold.change <- ((specific[,3] + specific[,4]) / specific[,3])
  specific <- specific[,-c(3,4)]

  return(specific)
}
```

Figure 8. Logbook excerpt of probesets, reference signals, and function (right) used to perform mDBR PSEA. Sample groups were defined as VTA = 0, SNpc = 1. Linear models constructed for each gene were filtered for models with an adjusted $R^2 > 0.50$, $p < 0.05$ for all regressors, an intercept less than 75% of the average expression, and a positive coefficient for the reference signal. The function, PSEA_cell_type_specific_mDBR, returns a data frame containing all genes exhibiting a significant population-specific fold change in the SNpc relative to the VTA.

```
Call:
lm(formula = AHBA_expr_SNpc_VTA_tranposed[, i] ~ micro_reference_mDBR +
    micro_difference_mDBR)

Residuals:
     Min       1Q   Median       3Q      Max
-0.76121 -0.32600  0.00389  0.26322  0.50903

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            6.7094     0.9236   7.264 6.33e-06 ***
micro_reference_mDBR   3.7055     0.9033   4.102  0.00125 **
micro_difference_mDBR  1.8360     0.2007   9.147 5.01e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3949 on 13 degrees of freedom
Multiple R-squared:  0.8784,    Adjusted R-squared:  0.8597
F-statistic: 46.94 on 2 and 13 DF,  p-value: 1.129e-06
```

```
Call:
lm(formula = AHBA_expr_SNpc_VTA_tranposed[, i] ~ micro_reference_mDBR +
    micro_difference_mDBR)

Residuals:
     Min       1Q   Median       3Q      Max
-0.52244 -0.16718 -0.01092  0.24633  0.37372

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)            7.2525     0.6652  10.902 6.55e-08 ***
micro_reference_mDBR   1.9417     0.6506   2.984   0.0105 *
micro_difference_mDBR  1.0499     0.1446   7.262 6.34e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2844 on 13 degrees of freedom
Multiple R-squared:  0.8166,    Adjusted R-squared:  0.7883
F-statistic: 28.93 on 2 and 13 DF,  p-value: 1.632e-05
```

Figure 9. Logbook excerpt of the summary output from R of microglia-based regression models for SNCA (top) and PINK1 (bottom) from the population-specific expression analysis.