



Inhibitor of DNA Binding Protein 3 (ID3) and Nuclear Respiratory Factor 1 (NRF1) Mediated Transcriptional Gene Signatures are Associated with the Severity of Cerebral Amyloid Angiopathy

Christian Michael Perez¹ · Zhenghua Gong² · Changwon Yoo² · Deodutta Roy¹ · Alok Deoraj¹ · Quentin Felty¹

Received: 23 March 2023 / Accepted: 25 July 2023 / Published online: 5 September 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Cerebral amyloid angiopathy (CAA) is a degenerative vasculopathy. We have previously shown that transcription regulating proteins- inhibitor of DNA binding protein 3 (ID3) and the nuclear respiratory factor 1 (NRF1) contribute to vascular dysregulation. In this study, we have identified sex specific ID3 and NRF1-mediated gene networks in CAA patients diagnosed with Alzheimer's Disease (AD). High expression of ID3 mRNA coupled with low NRF1 mRNA levels was observed in the temporal cortex of men and women CAA patients. Low NRF1 mRNA expression in the temporal cortex was found in men with severe CAA. High ID3 expression was found in women with the genetic risk factor APOE4. Low NRF1 expression was also associated with APOE4 in women with CAA. Genome wide transcriptional activity of both ID3 and NRF1 paralleled their mRNA expression levels. Sex specific differences in transcriptional gene signatures of both ID3 and NRF1 were observed. These findings were further corroborated by Bayesian machine learning and the GeNIe simulation models. Dynamic machine learning using a Monte Carlo Markov Chain (MCMC) gene ordering approach revealed that ID3 was associated with disease severity in women. NRF1 was associated with CAA and severity of this disease in men. These findings suggest that aberrant ID3 and NRF1 activity presumably plays a major role in the pathogenesis and severity of CAA. Further analyses of ID3- and NRF1-regulated molecular drivers of CAA may provide new targets for personalized medicine and/or prevention strategies against CAA.

Keywords Nuclear respiratory factor 1 · ID3 · Cerebral amyloid angiopathy · Vascular dementia · Transcriptomic signatures · Bayesian network

Introduction

Cerebral amyloid angiopathy (CAA) is a degenerative vasculopathy that leads to intracerebral hemorrhages and vascular related cognitive decline. CAA involves amyloid- β protein deposition in the leptomeningeal and cortical blood vessels. The buildup of amyloid- β in cerebral vessels contributes to the loss of smooth muscle cells, luminal narrowing, wall

thickening, and concentric wall splitting causing bleeding in the brain. Age is the most significant risk factor for CAA. Most cases of CAA are sporadic occurring over the age of 55, but rare familial cases of CAA also occur in younger patients from mutations in the amyloid precursor protein (APP) gene. The prevalence of moderate to severe CAA in the population is estimated to be 2.3% in 65- to 74-year-olds, 8% in 75- to 84-year-olds, and 12.1% in over 85-year-olds [1]. CAA is a significant cause of lobar intracerebral hemorrhages in the elderly [2]. Intracerebral hemorrhages account for 10–15% of total stroke cases in the U.S. each year with a devastating mortality rate of 54% at 1 year [3, 4]. In addition to stroke, CAA has been reported to influence cognitive decline in Alzheimer's Disease (AD) patients [5]. Currently there are no effective prevention strategies or disease modifying treatments for CAA.

✉ Quentin Felty
feltyq@fiu.edu

¹ Department of Environmental Health Sciences, Robert Stempel College of Public Health and Social Work, Florida International University, Miami, FL, USA

² Department of Biostatistics, Robert Stempel College of Public Health and Social Work, Florida International University, Miami, FL, USA

The APOE4 genotype has been shown to increase the accumulation of amyloid- β in the brain by inhibiting its clearance [6]. In CAA patients, carriers of APOE4 were shown to have greater amyloid accumulation in the cerebral vasculature [7]. APOE4 has also been associated with increased risk of AD in women [8]. Regardless of APOE4 genotype men have been shown to have more severe CAA than women [9]. Our previous studies have shown that redox sensitive transcription regulating proteins- inhibitor of DNA binding protein 3 (ID3) and the nuclear respiratory factor 1 (NRF1) contribute to vascular dysregulation [10–13]. Experiments have shown that amyloid- β protein induces inhibitor of DNA binding protein (ID) expression in rat cortical cells [14]. ID proteins have recently been implicated in the progression of AD [15] but their role in other amyloid related diseases such as CAA is not known. NRF1 transcription factor targets many genes associated with neurodegenerative disease [16]. The APOE4 gene contains a DNA binding motif for the NRF1 transcription factor [17].

Despite tremendous progress in understanding how amyloid- β protein and APOE4 contribute to CAA, gaps remain in understanding the molecular drivers of this small vessel disease. In this study, we have examined ID3 and NRF1-mediated networks in CAA patients. Using an RNA-seq case-control approach of CAA patients and AD patients, we examined ID3 and NRF1 mRNA levels associated with APOE4 genotype and CAA severity, their genome wide transcriptional activity, and sex specific differences in transcriptional gene signatures of both ID3 and NRF1. Causal Bayesian networks of ID3 and NRF1 were generated from published data using machine learning [18, 19] to determine their causal influence on CAA by a gain and loss of function approach [20]. Furthermore, we analyzed causal structures of ID3 and NRF1 networks to identify signatures involved in development of CAA by Markov Chain Monte Carlo (MCMC)-based gene order [21]. In summary, this study focuses on understanding sex specific differences associated with the transcriptional landscape of ID3 and NRF1 regulated genes to discover molecular drivers of CAA for personalized medicine and/or prevention strategies to address the small vessel disease that currently has no cure.

Results

High ID3 Expression Coupled with Low NRF1 Expression Levels are Associated with CAA

To evaluate whether ID3 and NRF1 are associated with CAA, we used RNA-seq data from the AD Knowledge Portal. The case-control approach was used to determine the association between the expression of ID3 and NRF1 in CAA by analyzing RNA-seq data. CAA favors the deposition of amyloid- β

protein in the temporal cortex and cerebellum while the hippocampus is often spared or less affected [22]. Our study was focused on the temporal cortex and the cerebellum because these were the only brain regions available with RNA-Seq data from CAA patients from the AD Knowledge Portal Synapse Repository. CAA patients had an average age of 81 yrs and were neuropathologically diagnosed with Alzheimer's disease (AD) (Table 17). Controls were composed of individuals with an average age of 82 yrs (Table 18). Differential gene expression (DEG) using limma and TMM normalization was conducted between these groups with significant genes determined by Benjamin-Hochberg FDR t-test ($p < 0.05$). Our analysis showed a significantly higher mean expression of ID3 mRNA in both men and women with CAA (Fig. 1A). CAA has been reported to be an overlapping pathology in AD patients. To determine if higher ID3 mRNA expression was specific to CAA, we also examined AD cases (Table 19). Our findings also showed higher ID3 mRNA in men and women AD patients. Because of the overlap in function between ID family proteins, we extended our efforts by analyzing ID1 mRNA. ID1 mean mRNA levels showed the same trend as ID3 with significantly high expression in CAA patients (Fig. 1B), but in the AD only cases ID1 expression was only higher in males.

We also examined NRF1 mRNA levels in CAA patients. As shown in Fig. 1C, there was a significantly lower level of NRF1 mRNA in men and women CAA and AD patients. CAA has also been shown to occur in the cerebellum [23]. Therefore, we determined mRNA levels of ID proteins and NRF1 in the cerebellum of CAA patients. As shown in Fig. 2A, ID3 mRNA levels were significantly higher in the cerebellum of CAA patients. AD cases also showed higher ID3 mRNA level compared to controls (Fig. 2A). Low NRF1 mRNA level in the cerebellum was significantly lower in men with AD but this was not observed in CAA (Fig. 2C). In summary, high ID3 expression coupled with low NRF1 expression levels in the temporal cortex are associated with men and women CAA patients. High ID3 and low NRF1 gene expression in the temporal cortex was also observed in AD patients.

In order to discern whether the effects on ID3 and NRF1 levels were CAA specific, we measured their levels in brain microvessels from AD patients. We analyzed the microarray data file GSE45596 from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO). Raw intensity values for the dataset were normalized according to log2 ratios and then analyzed using limma differential expression analysis. As shown in Fig. 3, the median ID3 and ID1 mRNA levels were higher in microvessels of AD cases and NRF1 mRNA levels were lower, however, they were not statistically significant (Fig. 3C). In summary, expression of ID3, ID1, and NRF1 mRNA levels in brain blood vessels from AD patients were similar to whole tissue RNA-seq levels of these genes in CAA and AD patients.

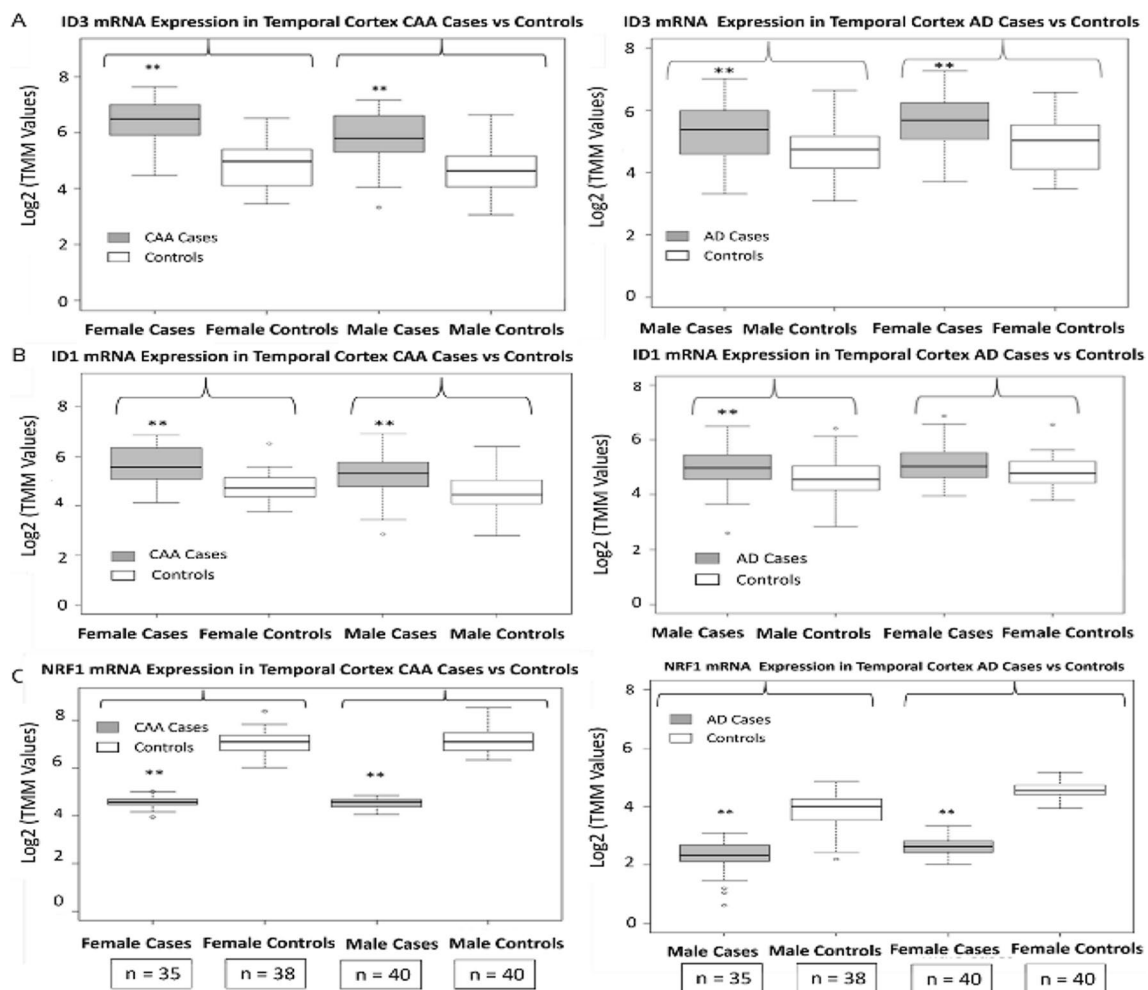


Fig. 1 Box plots of ID3, ID1, and NRF1 gene expression in the temporal cortex. High ID3 coupled with low NRF1 mRNA expression in the temporal cortex is associated with CAA compared to controls. **A** ID3 mRNA expression is higher in CAA cases vs controls in males ($p=1.38 \times 10^{-12}$) and females ($p=5.15 \times 10^{-8}$) according to two-tail t-test comparison. In AD Cases vs controls, ID3 mRNA levels were significantly higher in males ($p\text{-value}=0.001$) and females ($p=4.48 \times 10^{-5}$). **B** ID1 mRNA expression was significantly higher in CAA cases vs controls in males ($p\text{-value}=0.001$) and females ($p\text{-value}=6.11 \times 10^{-6}$). Low ID1 mRNA expres-

sion was significantly associated with male AD cases vs controls ($p\text{-value}=0.037$). **C** Low NRF1 mRNA level was significantly associated with CAA cases vs controls in males ($p=2.68 \times 10^{-41}$) and females ($p\text{-value}=4.484 \times 10^{-5}$). AD cases were significantly associated with lower NRF1 mRNA level in males ($p\text{-value}=1.458 \times 10^{-3}$) and females ($p\text{-value}=1.708 \times 10^{-3}$). The black line in the box plot represents the median value of gene expression. Asterisks indicate a significant difference between the mean expression level of cases vs controls according to t-test ($p\text{-value}<0.05$)

Low NRF1 Coupled with High ID3 Expression Correlated with CAA Severity in Temporal Cortex of Men

To determine whether ID3 and NRF1 expression levels were altered by CAA severity, we compared severe to non-severe cases. CAA severity scores were determined by thioflavin-S staining of amyloid- β and scored on a scale from 0 to 4. In our analysis, cases of CAA with a score of 2 or greater were included as severe cases. Severe cases represent strong circumferential amyloid deposition or extravasation of amyloid deposition in multiple cortical and leptomeningeal vessels.

Non-severe CAA cases were selected with a score of less than 2 representing scattered amyloid deposition in both leptomeningeal and cortical vessels. In the temporal cortex, ID3 expression was not associated with men or women with severe CAA (Fig. 4A). Low NRF1 expression was significantly associated in men with severe CAA (Fig. 4B).

We further determined if ID3 and NRF1 mRNA expression levels were altered by CAA severity in the cerebellum. Both men and women with severe CAA showed significantly higher ID3 mRNA expression (Fig. 5A). NRF1 expression was significantly downregulated in both men and women with severe CAA in the cerebellum. In summary, high ID3 and low NRF1 expression

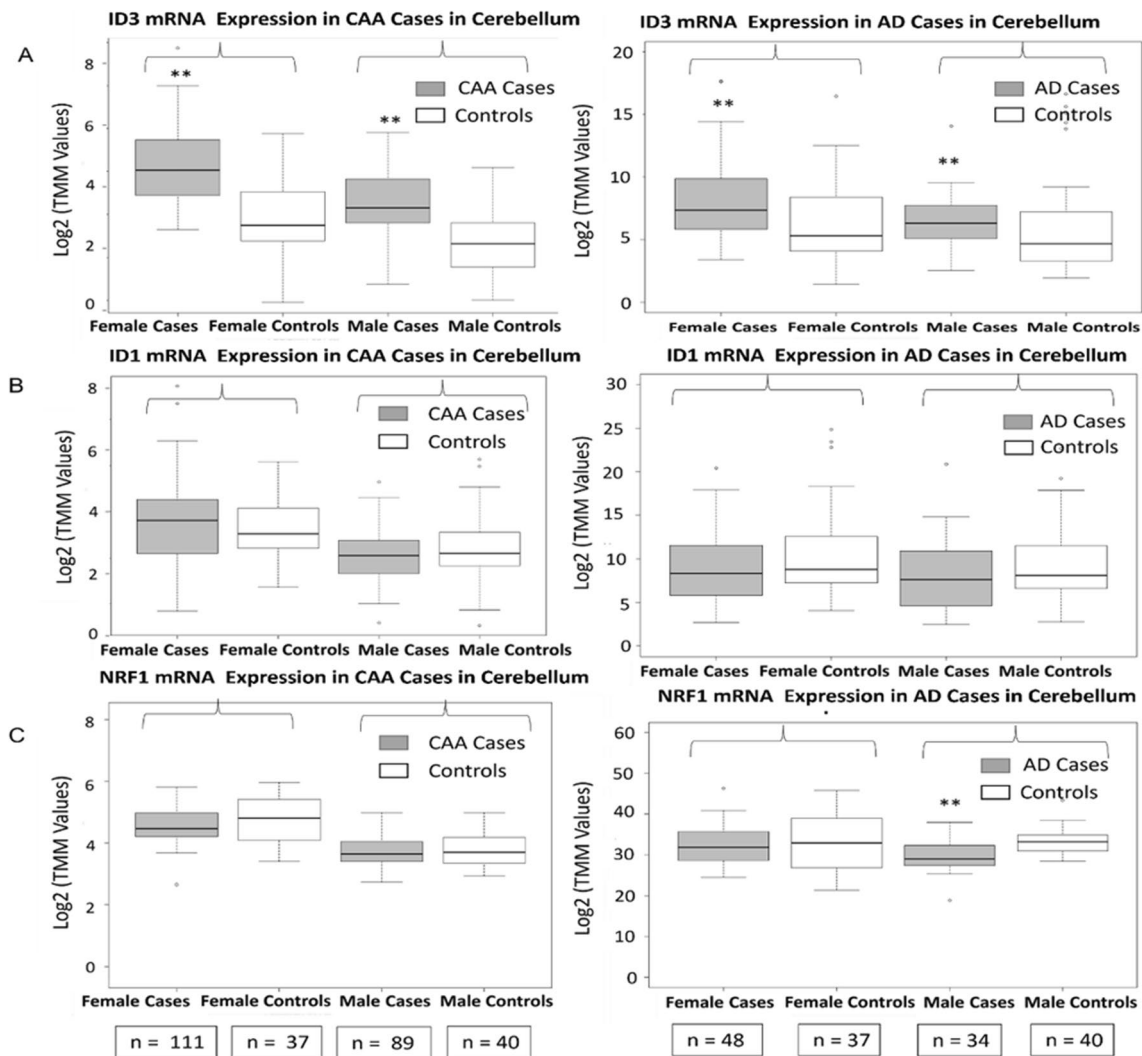


Fig. 2 Box plots of ID3, ID1, and NRF1 gene expression in the cerebellum. High ID3 mRNA expression in the cerebellum is associated with CAA cases. **A** ID3 mRNA expression is higher in CAA cases vs controls in males ($p=1.14 \times 10^{-6}$) and females ($p=5.68 \times 10^{-7}$) according to two-tail t-test comparison. In AD cases vs controls, ID3 mRNA expression was significantly higher in females ($p=4.32 \times 10^{-2}$) and in males. **B** ID1 mRNA expression was not associated with

either CAA or AD cases. **C** NRF1 mRNA expression in AD cases vs controls was significantly downregulated in males ($p\text{-value}=0.003$), but not in AD female cases. The black line in the box plot represents the median value of gene expression. Asterisks indicate a significant difference between the mean expression level of cases vs controls according to t-test ($p\text{-value} < 0.05$)

were associated with severity in the cerebellum for both men and women. As described previously, men are diagnosed more frequently with severe CAA compared to women. Cerebral arterioles are reported to show significant CAA-dependent loss of α -smooth muscle actin at Braak stage 5 and 6 in AD subjects [24]. So we further examined the effect of CAA severity according to Braak stage in the temporal cortex. As shown in Fig. 6A, significantly higher ID3 mean mRNA expression was found in men with severe cases of CAA and a Braak stage of 5/6 compared to non-severe cases. In summary, high ID3 expression in the temporal cortex was associated with men who had the combination of severe CAA and a Braak stage of 5/6. Low NRF1 expression was also associated with men who had severe CAA.

ID3 and NRF1 Expression by APOE4 Genotype

Since APOE4 with the NRF1 DNA element are considered as risk factors of AD, we examined ID3 and NRF1 expression in different APOE4 genotypes. Subjects with one or more alleles of APOE4 were defined to be carriers. As shown in Fig. 7B, high ID3 expression in the temporal cortex was significantly associated with women carrying APOE4. Women with APOE4 showed significantly lower NRF1 mRNA expression (Fig. 7C). We further determined ID3 and NRF1 mRNA expression by APOE4 genotype in the cerebellum. ID3 expression was significantly upregulated in both men and women

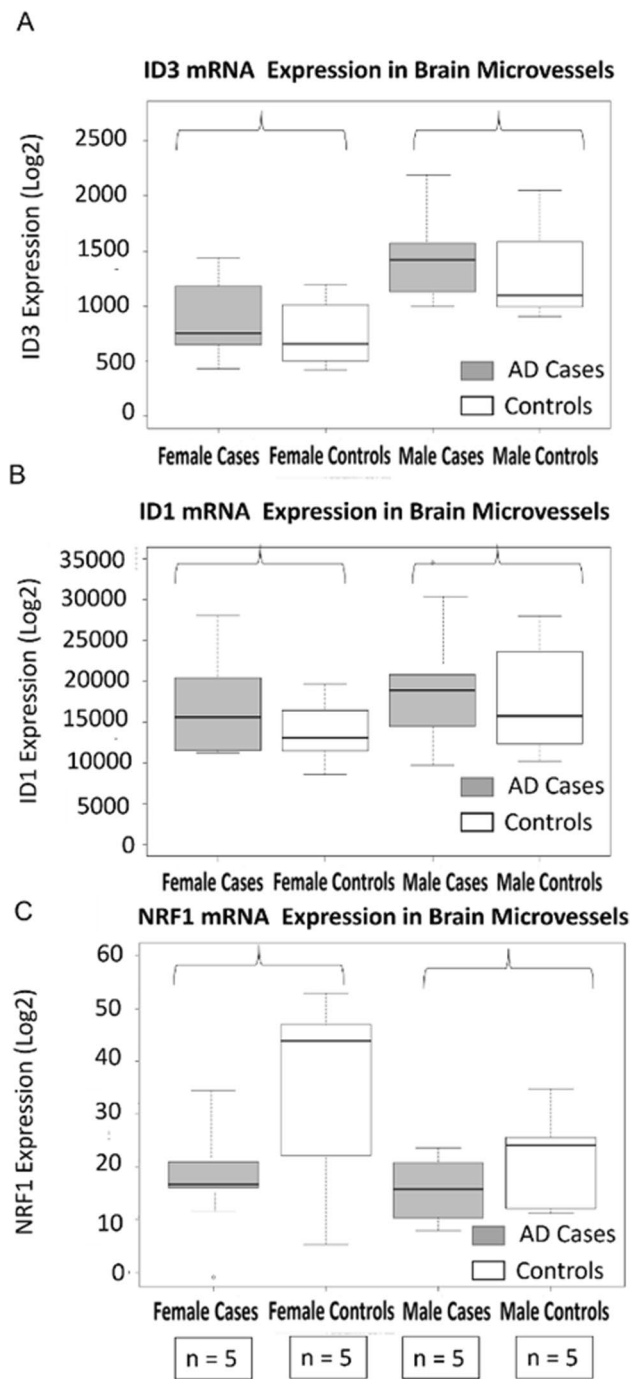


Fig. 3 Box plots of ID3, ID1, and NRF1 gene expression in brain microvessels. **A** Median ID3 mRNA expression levels are higher in blood vessels from AD cases compared to controls. **B** Median ID1 mRNA expression levels are higher in AD cases vs controls. **C** Median NRF1 mRNA expression level is lower in AD cases vs controls. GSE45596 microarray data file of brain microvessels from AD patients ($n=5$ female, $n=5$ male) vs controls ($n=5$ female, $n=5$ male). The black line in the box plot represents the median value of gene expression

who carried APOE4 (Fig. 8A). In summary, sex-specific association of low NRF1 expression and APOE4 in the temporal cortex was found in women. We also show high ID3 expression associated with APOE4 in the temporal cortex of women. These findings in the temporal cortex are consistent with the higher risk of AD in women with APOE4. However, this should be interpreted cautiously because sample size was low as well as the appropriate controls were lacking.

Genome wide Transcriptional Activity of ID3 and NRF1

To determine if high ID3 and low NRF1 mRNA levels observed in CAA patients impacted genome wide transcriptional activity, we conducted a transcriptional activity assay using LRPATH logistic regression analysis. Using LRPATH we examined the total transcriptional activity of ID3 and NRF1 based on the expression of their target genes for discovery of CAA-associated transcriptional activity. Transcriptional activity of ID3 and NRF1 was determined from the cumulative statistically significant differential expression of hundreds of target genes. The results that follow are all visualized as a heatmap shown in Fig. 9.

ID3 transcriptional activity was significantly upregulated in the temporal cortex of women ($p=0.004$) and men ($p=0.010$) CAA cases compared to controls shown in Fig. 9. In men with CAA, we found that 67% (8508/9438) of all known ID3 target genes have significantly altered mRNA expression, with 38% (2431/6348) of these being downregulated, and 62% (3917/6348) of these being upregulated in the temporal cortex. In women with CAA, we found that 90.15% (8508/9438) of all known ID3 target genes have significantly altered mRNA expression, with 46.34% (3943/8508) of these being downregulated in CAA, and 54.36% (4057/8508) of these being up-regulated in the temporal cortex.

Our analysis showed a downregulation of NRF1 transcriptional activity in the temporal cortex of men and women patients with CAA. We showed that 63% (7099/11280) of all known NRF1 target genes have significantly altered mRNA expression, with 62% (4407/7099) of these being downregulated in men with CAA, and 38% (2692/7099) of these being upregulated. In women with CAA, we found that 62% (6942/11280) of all known NRF1 target genes have significantly altered mRNA expression in cases, with 58% (4027/6942) of these being downregulated in cases, and 42% (2915/6942) of these being upregulated. In summary, genome-wide transcriptional activity of ID3 and NRF1 in CAA patients correlated with high ID3 and low NRF1 mRNA levels in CAA patients.

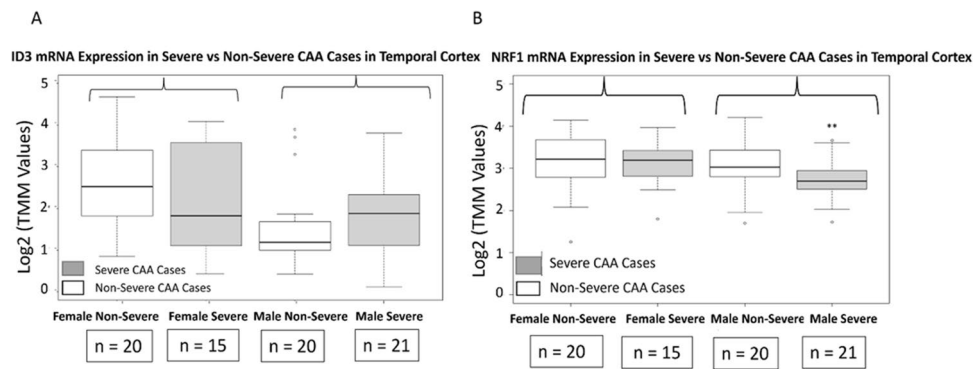


Fig. 4 Box plots of ID3 and NRF1 gene expression in the temporal cortex of severe vs non-severe CAA cases. **A** ID3 mRNA levels across severe male CAA cases ($n=21$) vs male non-severe CAA cases ($n=20$) and the distribution of mRNA expression in severe female CAA as compared to non-severe CAA were not significantly different according to two-tail t-test comparison. **B** NRF1 mRNA expression was significantly lower in male severe CAA cases ($n=21$)

($p=0.048$) vs male non-severe CAA cases ($n=19$), but not in female severe CAA cases ($n=15$) vs female non-severe CAA cases ($n=20$) according to t-test comparison. The black line in the box plot represents the median value of gene expression. Asterisks indicate a significant difference between the mean expression level of severe cases vs non-severe cases according to t-test (p -value < 0.05)

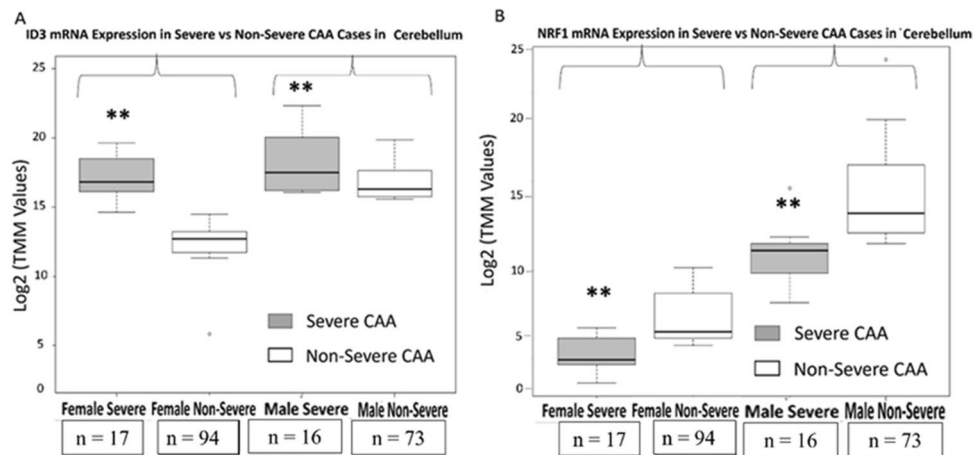


Fig. 5 Box plots of ID3 and NRF1 gene expression in the cerebellum of severe vs non-severe CAA cases. **A** ID3 mRNA expression levels across severe male CAA cases ($n=16$) ($p=0.0208$) vs male non-severe CAA cases ($n=73$) were not significantly upregulated, however the distribution of mRNA expression in severe female CAA ($n=17$) as compared to non-severe female CAA cases ($n=94$) ($p=0.0009$) was significantly upregulated according to two-tail t-test comparison. **B** NRF1 mRNA expression was significantly lower in

male severe CAA cases ($n=16$) ($p=0.037$) vs male non-severe CAA cases ($n=73$) and in expression of female severe CAA cases ($n=17$) vs female non-severe CAA cases ($n=94$) ($p=0.007$) according to two-tail t-test comparison. Severe CAA is a severity score above 2. The black line in the box plot represents the median value of gene expression. Asterisks indicate a significant difference between the mean expression level of severe cases vs non-severe cases according to t-test (p -value < 0.05)

We further validated genome wide transcriptional activity of ID3 and NRF1 in the temporal cortex of AD patients. In men with AD, we found that 67% (8508/9438) of all known ID3 target genes have significantly altered mRNA expression with 38% (2431/6348) of these being downregulated and 62% (3917/6348) of these being upregulated. In women with AD, we found that 90.15% (8508/9438) of all known ID3 target genes have significantly altered mRNA expression, with 46.34% (3943/8508) of these being downregulated, and 54.36% (4057/8508) of these being up-regulated.

For NRF1 transcriptional activity in men AD patients, we found that 61.6% (6954/11280) of all known NRF1 target genes have significantly altered mRNA expression with 45% (3181/6954) being upregulated and 54% (3773/6954) being downregulated. Women with AD showed 36.9% (4171/11280) of all known NRF1 target genes to have significantly altered mRNA expression, with 66.3% (2767/4171) being upregulated and 34.7% (1404/4171) being downregulated. Results for ID3 and NRF1 transcriptional activity by pathway analysis can be found in (Tables 1, 2, 3, 4, 5, and 6).

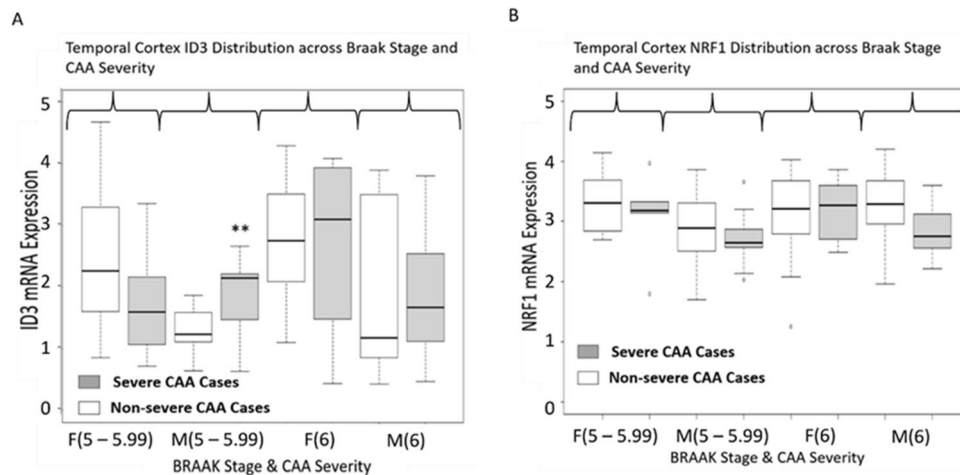


Fig. 6 Box plots of ID3 and NRF1 gene expression by Braak Stage and CAA severity in temporal cortex of CAA cases. ID3 mRNA expression is higher in men with severe cases of CAA ($n=37$) and a Braak stage of 5/6 in the temporal cortex. **A** ID3 expression in severe cases of CAA compared to non-severe cases with Braak score of 5 or higher. ID3 expression is significantly higher (p -value=0.0256) in males ($n=37$) with severe cases of CAA and with a Braak stage

of 5 to 5.99. **B** NRF1 expression in severe cases of CAA compared to non-severe cases with Braak of 5 or higher. The black line in the box plot represents the median value of gene expression. Asterisk indicate significant difference between the mean expression level of severe CAA vs non-severe CAA cases across a braak stage range of (5 – 5.99 or ≥ 6) according to t-test (p -value < 0.05)

In brief, genome wide transcriptional activity of both ID3 and NRF1 paralleled their mRNA expression levels. Gene pathways significantly downregulated in NRF1 total activity included endothelial cell signature, VEGF signaling, endothelial cell angiogenesis and vasculogenesis, and Wnt Signaling Pathway. Downregulation of angiogenesis and VEGF signaling could explain how NRF1 mediates injury of cerebral blood vessels in CAA.

ID3 Bayesian Network for CAA in Women

On the basis of high ID3 mRNA expression in the temporal cortex of women that had CAA and the association of ID3 with APOE4 in women with AD, we generated probabilistic graphical models to learn sex specific molecular drivers of CAA. In this study, nodes consisted of RNA-seq CAA case and control expression data for ID3 target genes from women and clinical variables (CAA disease, APOE4 genotype, Thal phase, and Braak stage). To learn the CBNs, we ran independent runs of BANJO to identify Markov blanket MB) genes. The first-degree MB genes of variable CAA in the CBN (denoted as MB [CAA]) was defined as the set of variables that represent the direct causes (parents) of the direct effects (children) of CAA. The scoring metric used is called Bayesian Dirichlet equivalence (BDe). Among the 12 best runs for the ID3 women CBN, structures with the best BDe score (–16,204.98) were selected for comparison of CAA MB genes amongst the best consensus structures with 318 variables and 73 patients (Fig. 10). During the validation runs for MB genes for CAA, CBN learning comprising

the nine CBNs, the best scoring CBN with a BDe score (–4010.11) contained 75 variables with 73 patients (Fig. 11). Thal phase node was the plausible direct cause (parent) of the CAA node. Thal A β phase predicts A β deposition in cortical and subcortical areas of brain tissue increasing risk of having CAA [25]. The node for ID3 was directly connected to CAA. This indicates that CAA was a result of ID3. Furthermore, CAA was the plausible direct cause of the severity node. This indicates that severity of neuropathological and cognitive conditions was a result of CAA status at the time of death. CAA was the plausible direct cause of 39 genes in total (Table 7), indicating the importance of disease status to the expression of these ID3 targeted genes. The influence scores between arcs (connections) in the CBN for ID3 was calculated with the BANJO algorithm. Using a correlation score of ± 0.70 as our threshold, we found 13 CAA MB variables that were used in the order scoring algorithm to determine the causal gene order signatures.

NRF1 Bayesian Network for CAA in Men

Since low NRF1 mRNA expression in the temporal cortex was specific to men with severe CAA, we used probabilistic graphical models to learn if NRF1 was a sex specific driver of CAA. Among the 12 best runs for the NRF1 male CBN, structures with the best BDe score (–14,349.32) were selected for comparison of CAA MB genes with 289 variables and 81 patients (Fig. 12). During the validation runs for MB genes for CAA, CBN learning comprising the nine CBNs, the best scoring CBN was with a BDe

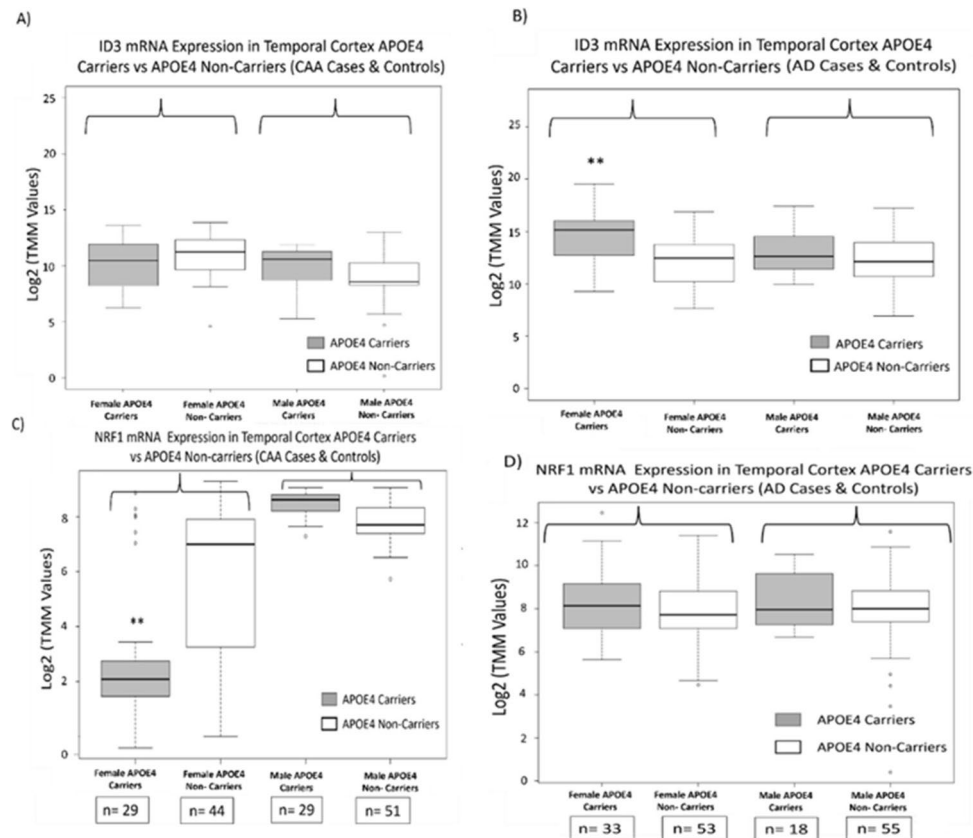


Fig. 7 Box plots of ID3 and NRF1 gene expression in the temporal cortex of APOE4 carriers. Subjects consisting of both cases and controls with one or more alleles of APOE4 were defined carriers. **A** ID3 expression by APOE4 genotype in a population of CAA cases and controls. **B** ID3 expression by APOE4 genotype in a population of cases with AD and controls. ID3 mRNA expression was significantly higher (p -value = 1.72×10^{-5}) in female APOE4 carriers compared to non-carriers. **C** NRF1 expression by APOE4 genotype in a

population of CAA cases and controls. NRF1 mRNA expression was significantly lower (p -value = 1.24×10^{-6}) in women APOE4 carriers when compared to non-carriers. **D** NRF1 expression by APOE4 genotype in a population of AD cases and controls. The black line in the box plot represents the median value of gene expression. Asterisk indicates significant difference between the mean expression level of target gene in APOE4 carriers vs non-carriers consisting of both cases and controls according to t-test (p -value < 0.05)

score (-5167.7363) containing 93 variables and 81 patients (Fig. 13). The highest scoring NRF1 representative network consisted of two parent genes (SERHL, PXMP4) that represented the plausible direct causes of CAA in the network. SERHL is a serine hydrolase like pseudo-gene that is associated with vascular dysfunction [26, 27]. PXMP4 is a protein coding gene that is associated with cerebrohepatorenal syndrome and peroxisome biogenesis disorders responsible for processes that promote formation of vascular diseases [28, 29]. CAA node was the direct cause of the Thal phase node. Thal A β phase predicts A β deposition in cortical and subcortical areas of CAA brain tissue which supports the biological plausibility of the network. A direct connection between NRF1, Thal, and CAA was identified in the CBN. CAA was the direct cause of 51 genes and microRNAs in total (Table 8), indicating the importance of disease status to the expression of these NRF1 target genes.

ID3 Markov Blanket Causal Influence Analysis

After generating the ID3 CBN structures, the second objective was to learn the probable contribution of ID3 to CAA. A sensitivity analysis using GeNIe was conducted on experimentally modified ID3 expression status for upregulation, downregulation, and normal expression. We observed an increase in marginal probability of CAA from 48% to a 91% risk following changing ID3 expression to an upregulated state. In a state of downregulation of ID3, we observed a decrease in marginal probability of CAA from 48 to 8% risk validating that ID3 overexpression drives CAA risk. In a state of normal expression for ID3, the marginal probability of CAA was at a state of 48% and increased to a state of 49% when expression was modified. In conjunction with Bayes Theorem, the lifetime risk of CAA was calculated for single gene, gene pairs, and multiple genes. Our control group age ranges from 79–84 yrs old and the prevalence of CAA in this

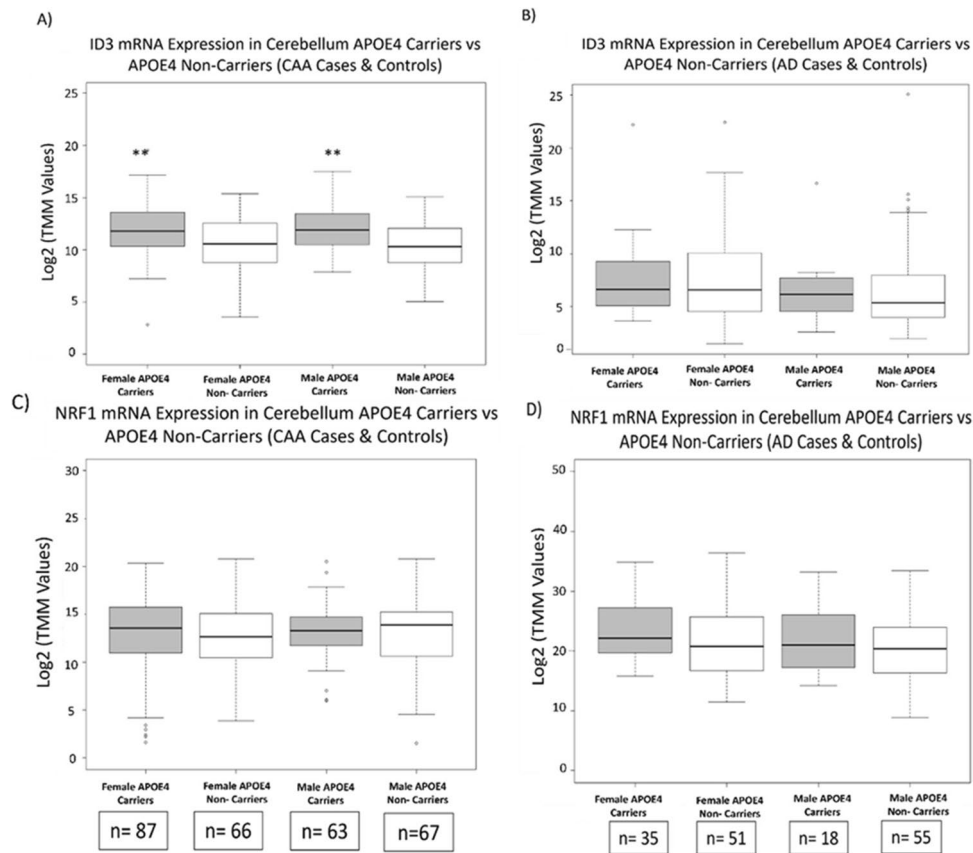


Fig. 8 Box plots of ID3 and NRF1 gene expression in the cerebellum of APOE4 carriers. Subjects consisting of both cases and controls with one or more alleles of APOE4 were defined carriers. **A** Higher ID3 expression is associated with APOE4 genotype in a population of CAA cases and controls. ID3 mRNA expression was significantly higher (p -value = 1.44×10^{-3}) in female APOE4 carriers when compared to non-carriers. ID3 expression was also higher (p -value = 1.73×10^{-5}) in male APOE4 carriers. **B** ID3 expression by

APOE4 genotype in a population of cases with AD and controls. **C** NRF1 expression by APOE4 genotype in a population of CAA cases and controls. **D** NRF1 expression by APOE4 genotype in a population of AD cases and controls. The black line in the box plot represents the median value of gene expression. Asterisk indicates significant difference between the mean expression level of the target gene in APOE4 carriers vs non-carriers consisting of both cases and controls according to t-test (p -value < 0.05)

age group has reported to be 8% [1]. Therefore, we used this prevalence for calculating the lifetime risk. The ID3 target genes and gene combinations that affect lifetime risk CAA are shown in Table 9.

Several gene candidates contributed more significantly to lifetime CAA risk or P (CAA). 8% which we used as a baseline risk for CAA in our population diagnosed with AD (Calculating modified expression levels and their effect on the probability of evidence for CAA we were able to determine the lifetime risk of CAA shown in Table. Overexpression of ID3 itself led to a significantly higher probability of CAA increasing lifetime CAA risk > sixfold. In combination with other gene expression patterns for our Markov blanket genes determined by our model the risk for CAA increased exponentially in concert with ID3 overexpression. Overexpression of ATG10, SLC5A2, UTS2, DNAJC25, and U2AF1 in combination with ID3 overexpression resulted in the lifetime risk of CAA given the data increasing by

> 12-fold. Underexpression of HIC1 in combination with overexpression of ID3 resulted in the lifetime risk of CAA increasing by 12.24 -fold (Fig. 14).

ID3 Markov Blanket Combinatorial Relative Risk Analysis

The PREDICT combinatorial analysis was used to confirm results from GeNIe and to calculate risk ratios for sets of gene combinations involved in contributing to the risk of CAA according to GeNIe results. We used the SMILE and C++ based combinatorial analysis program PREDICT to determine expression patterns most likely to contribute to CAA risk in women using the MB genes in the ID3 representative network. Utilizing the ID3 representative network GeNIe's 'learn' parameters calculated the probability of evidence for CAA likelihood given the modified expression levels of the 16 genes (including NRF1 and ID3) associated

Fig. 9 Transcriptional activity of ID3 and NRF1 in temporal cortex in CAA and AD. In the heatmap above, NRF1 activity is strongly downregulated and ID3 activity is upregulated. Other transcription factors shown above represent the most significant probability of transcriptional activity across both diseases in the temporal cortex

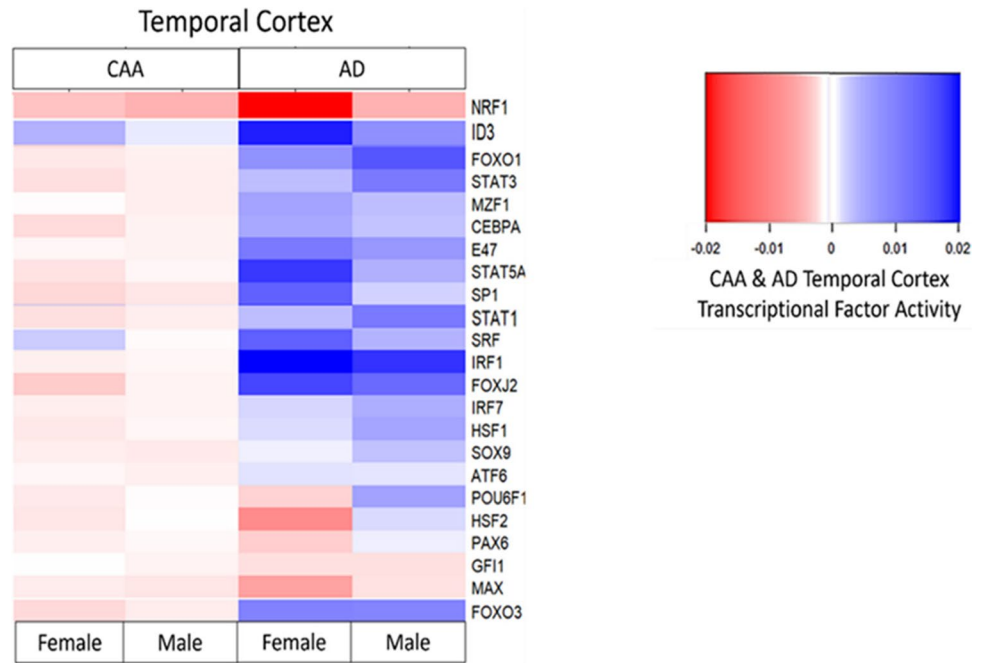


Table 1 NRF1 and ID3 pathway activity in female CAA cases vs controls at Temporal Cortex

	Pathway activity	Total target gene no	No. significant genes	Coefficient	Odds ratio	P-Value	Total activity
NRF1	NRF1 Total Activity	11,280	6942	-0.02610	0.85	9.02E-07	Down*
	NRF1 Endothelial Signature	1937	142	-0.007	0.956	1.08E-04	Down*
	VEGF Signaling	56	29	-0.00866	0.947	0.0447	Down*
	Apoptosis	141	37	-0.00376	0.976	0.6789	Down
	Cell Cycle	42	7	0.0010	1.007	0.944	Up
	Cellular Senescence	49	17	0.0190	1.125	0.270	Up
	Epigenetic Chromatin Remodeling Factors	67	26	0.0055	1.034	0.146	Up
	Endothelial Cell Death	10	2	-0.0113	0.932	0.099	Down
	Endothelial Cell Growth & Proliferation	13	9	-0.014	1.1633	0.372	Down
	Endothelial Cell Angiogenesis & Vasculogenesis	42	22	-0.0154	0.90835	0.017	Down*
ID3	ID3 Total Activity	9438	2374	0.005476	1.0346	1.05E-10	Up*
	ID3 Endothelial Signature	2835	1535	0.004	1.021	2.03 E-11	Up*
	VEGF Signaling	53	15	6.48E-04	1.004	0.872	Up
	Apoptosis	208	65	0.0264	1.178	0.057	Up
	Cell Cycle	76	20	0.0150	1.098	0.525	Up
	Cellular Senescence	67	21	0.0050	1.032	0.783	Up
	Epigenetic Chromatin Remodeling Factors	56	21	-0.0063	0.962	0.099	Down
	Endothelial Cell Death	13	9	0.0031	1.019	0.640	Up
	Endothelial Cell Growth & Proliferation	14	7	-0.004	1.1364	0.453	Down
	Endothelial Angiogenesis & Vasculogenesis	46	20	-0.0055	0.965	0.3670	Down

Significant pathways (p -value ≤ 0.05 , Fishers Exact Test) denoted by asterisks

with the causal Bayesian network for CAA in women. To validate the key ID3 representative network Markov blanket genes as causal signature genes for CAA women, we evaluated their predictive capability to distinguish between

normal and CAA cases. Genie's 'learn parameters' function analysis of the 16 genes (including ID3) associated with the CAA network showed that 5 Markov blanket genes (ID3, ATG10, SLC5A2, U2AF1, and UTS2) consistently

Table 2 NRF1 and ID3 pathway analysis in male CAA cases vs controls at Temporal Cortex

	Pathway activity	Total target gene no	No. significant genes	Coefficient	Odds ratio	P-Value	Total activity
NRF1	NRF1 Total Activity	11,280	7099	−0.09220	0.56	1.55E-26	Down*
	NRF1	1937	142	−0.006	0.96	5.04E-04	Down*
	Endothelial Signature						
	VEGF Signaling	56	17	0.04272	1.340	0.01104	Up*
	Apoptosis	141	49	−0.0148	0.912	0.129476	Down
	Cell Cycle	42	11	8.94865E-06	1.000012	0.9995	Up
	Cellular Senescence	49	11	0.0015	1.0096	0.9285	Up
	Epigenetic Chromatin Remodeling Factors	56	19	−0.006	0.9629	0.7135	Down
	Endothelial Cell Death	13	1	0.004	1.028	0.870	Up
	Endothelial Cell Growth & Proliferation	13	1	0.024	1.163	0.373	Up
ID3	Endothelial Cell Angiogenesis & Vasculogenesis	42	10	0.026	1.177	0.372	Up
	ID3 Total Activity	9438	2105	0.00307	1.304071	0.0103	Up*
	ID3 Endothelial Signature	1937	464	−0.0257	0.853	1.63E-08	Down*
	VEGF Signaling	53	11	0.011186	1.071988	0.437	Up
	Apoptosis	208	58	1.10E-02	1.070916	0.418	Up
	Cell Cycle	76	16	0.0030	1.0193	0.881	Up
	Cellular Senescence	67	20	−0.0055	0.9659	0.765	Down
	Epigenetic Chromatin Remodeling Factors	67	13	0.011	1.073	0.495	Up
	Endothelial Cell Death	10	3	0.056	1.422	0.149	Up
	Endothelial Cell Growth & Proliferation	14	1	0.021	1.136	0.454	Up
	Endothelial Angiogenesis & Vasculogenesis	46	9	−0.0250	0.8556	0.503	Down

Significant pathways (p -value ≤ 0.05 , Fishers Exact Test) denoted by asterisks

distinguished between female control and female CAA cases.

Table 10 summarizes the top 7 maximum relative risk (RR) of the minimum set of combination of gene expression patterns in the ID3 causal Bayesian network in women for determining CAA status. The likelihood of CAA is almost 100% in a patient with the expression pattern of (high) ID3 combined with ATG10 (high), SLC5A2 (high), U2AF1 (high), and UTS2 (high). However, a subject that has (low) ID3 expression combined with ATG10 (low or no change), SLC5A2 (low or no change), U2AF1 (low or no change), and UTS2 (low or no change) has almost zero probability of CAA. This discovery confirms the association of high ID3 overexpression combined with specific target genes to be most influential to CAA status, showing high probability of CAA in women.

NRF1 Markov Blanket Causal Influence Analysis

The sensitivity analysis was conducted on modified NRF1 expression status for downregulation, upregulation, and normal expression. We observed an increase in marginal probability of CAA from 50% to an 85% when changing NRF1 expression status to a downregulated state (Fig. 15). In a state of upregulation of NRF1, we observed a decrease in

marginal probability of having CAA from 50 to 24%. These results indicate that low NRF1 expression drives CAA risk. In a state of normal expression for NRF1, the marginal probability of developing CAA decreased slightly from 50 to 49%. In conjunction with Bayes Theorem, the lifetime risk of CAA was calculated for single gene, gene pairs, and multiple genes. The NRF1 target genes and gene combinations in the GeNIe modeler can be seen in Fig. 15.

Several gene candidates contributed more significantly to lifetime CAA risk or P (CAA) in the NRF1 representative network for CAA patients (Table 11). The baseline prevalence of CAA in patients is 8% in the population. Calculating modified expression levels and their effect on the probability of evidence for CAA, we were able to determine the lifetime risk of CAA. Downregulation of NRF1 itself led to a significantly higher probability of CAA increasing lifetime CAA risk > fourfold. In combination with other gene expression patterns for our Markov blanket genes determined by BANJO modeling the risk for CAA increased exponentially in concert with NRF1 downregulation. Overexpression of GDF9, SNORA33, ZNF135, SLC9B1, SNORA64, SATL1, SERHL, SNORA31, RPL39, MCM3AP-AS1, and EXOSC6 in combination with NRF1 downregulation resulted in the lifetime risk of CAA patients increasing by > ninefold over the baseline risk. The results are significant in that these

Table 3 ID3 & NRF1 pathway activity in male AD Cases vs Controls at Temporal Cortex

	Pathway activity	Total target gene no	No. significant genes	Coefficient	Odds ratio	P-Value	Total activity
NRF1	NRF1 Total Activity	11,280	326	−0.092	0.56	1.55 X 10 ^{−26}	Down*
	NRF1 Endothelial Signature	1937	200	−0.006	0.86	7.04E-08	Down*
	VEGF Signaling	56	10	0.061	1.46	0.408	Up
	Apoptosis	142	28	0.026	1.18	0.519	Up
	Cell Cycle	76	8	1.91E-05	1.001	0.999	Down
	Cell Senescence	67	15	0.036	1.25	0.555	Up
	Epigenetic Chromatin Remodeling Factors	67	15	0.204	3.555	0.009	Up*
	WNT Signaling Pathway	82	22	0.09802	1.8388	0.1713	Up
	Endothelial Cell Death	13	7	0.1888	3.2322	0.0512	Up
	Endothelial Cell Growth and Proliferation	13	7	0.2561	4.9119	0.0191	Up
	Endothelial Cell Angiogenesis & Vasculogenesis	39	19	0.0908	1.7581	0.3383	Up
	Endothelial Cell Cycle	28	3	−0.1224	0.4673	0.4286	Down
	Endothelial Cell Development	13	5	0.054	1.402	0.769	Up
	Endothelial Cell Movement	17	10	0.149	2.526	0.359	Up
ID3	ID3 Total Activity	9438	1130	0.09	1.707	4.61 X 10 ^{−16}	Up*
	ID3 Endothelial Signature	2619	731	0.04	1.876	3.08 E-24	Up*
	VEGF Signaling	53	10	0.031	1.21	0.673	Up
	Apoptosis	211	42	0.066	1.50	0.271	Up
	Cell Cycle	42	5	−0.048	0.741	0.572	Up
	Cell Senescence	49	12	0.010	1.06	0.879	Up
	Epigenetic Chromatin Remodeling Factors	56	4	−0.236	0.2313	0.004	Down*
	WNT Signaling Pathway	69	16	0.02768	1.1877	0.6687	Up
	Endothelial Cell Death	10	4	0.1397	2.3825	0.1558	Up
	Endothelial Cell Growth & Proliferation	14	6	0.1773	3.0098	0.0741	Up
	Endothelial Angiogenesis & Vasculogenesis	45	2	−0.3716	0.0993	0.0059	Down*
	Endothelial Cell Cycle	19	2	−4.98E-12	1	1	Down
	Endothelial Cell Development	14	0	−0.270	0.187	0.185	Down
	Endothelial Cell Movement	21	0	−0.224	0.249	0.290	Down

Significant pathways (p -value ≤ 0.05 , Fishers Exact Test) denoted by asterisks

genes may serve as potential drivers for the pathological processes that contribute to CAA. The inactivation of NRF1 combined with the inactivation of TRIM44 and GATC significantly increased lifetime risk of CAA by > 12-fold. Furthermore, we examined the combined effect of NRF1 inactivation and ID3 activation leading to increased CAA lifetime risk who are men. The combined effect significantly increased lifetime risk of CAA in by > tenfold validating the results of the genome wide transcriptional activity analysis conducted earlier.

NRF1 Markov Blanket Combinatorial Relative Risk Analysis

To further analyze the contribution of NRF1 Markov blanket causal genes to CAA risk, we analyzed the predictive capability of the 18 Markov blanket genes including NRF1 to discern their effects in normal and

CAA cases in men. We used the SMILE and C++ based combinatorial analysis program PREDICT to determine expression patterns most likely to contribute to CAA risk in men amongst the MB genes in the NRF1 representative network. Utilizing the NRF1 representative network GeNIe's 'learn' parameters calculated the probability of evidence for CAA likelihood given the modified expression levels of the 18 genes (including NRF1 and ID3) associated with the causal Bayesian network for CAA in men. Consistently, 5 Markov blanket gene targets (GATC, GDF9, TOMM6, SLC9B1, ZNF135) of NRF1 were found to contribute most to the likelihood of CAA in the PREDICT combinatorial analysis. In addition to the lifetime CAA risk calculated for the NRF1 target genes, the prediction accuracy to distinguish normal or CAA cases was alternatively verified by expression patterns of these combinations of genes.

Table 4 ID3 and NRF1 pathway activity in female AD cases vs Controls at Temporal Cortex

	Pathway activity	Total target gene no	No. significant genes	Coefficient	Odds ratio	<i>P</i> -Value	Total activity
NRF1	NRF1 Total Activity	11,280	2290	−0.026	0.850	9.018 E ^{−7}	Down*
	NRF1 Endothelial Signature	2227	475	−0.03918	0.784	5.22 E ^{−10}	Down*
	VEGF Signaling	56	11	0.0312	1.214	0.674	Up
	Apoptosis	142	28	0.0263	1.178	0.519	Up
	Cell Cycle	76	8	−0.04810	0.741	0.571	Down
	Cell Senescence	67	15	0.01009	1.065	0.879	Up
	Epigenetic Chromatin Remodeling Factors	67	15	0.20413	3.556	0.0094	Up*
	WNT Signaling Pathway	82	22	0.1209	2.12	0.0226	Up*
	Endothelial Cell Death	13	7	0.188	3.232	0.0122	Up*
	Endothelial Cell Growth and Proliferation	14	6	0.256	4.91	0.0191	Up*
	Endothelial Cell Angiogenesis & Vasculogenesis	39	19	0.09	1.758	0.338	Up
	Endothelial Cell Cycle	28	3	0.00000004	0.999	0.99	Down
	Endothelial Cell Development	13	5	0.054	0.271	0.769	Down
	Endothelial Cell Movement	17	10	0.149	2.53	0.359	Up
ID3	ID3 Total Activity	9384	1130	0.0432	1.307	1.052 E ^{−10}	Up*
	ID3 Endothelial Signature	2619	641	0.0614	1.465	1.08 E ^{−21}	Up*
	VEGF Signaling	53	15	0.00065	1.004	0.872	Up
	Apoptosis	208	78	−0.000041	0.995	0.990	Down
	Cell Cycle	42	10	0.01714	1.112	0.702	Up
	Cell Senescence	49	20	0.00490	1.030	0.904	Up
	Epigenetic Chromatin Remodeling Factors	56	6	−0.0714	0.641	0.093	Down
	WNT Signaling Pathway	69	23	0.0329	1.227	0.462	Up
	Endothelial Cell Death	10	4	0.0733	1.577	0.320	Up
	Endothelial Cell Growth & Proliferation	14	7	0.144	2.44	0.041	Up*
	Endothelial Angiogenesis & Vasculogenesis	45	1	−0.1059	0.517	0.174	Down
	Endothelial Cell Cycle	19	2	−0.0498	0.733	0.717	Down
	Endothelial Cell Development	13	0	−0.021	0.875	0.869	Down
	Endothelial Cell Movement	21	0	−0.166	0.356	0.267	Down

Significant pathways (*p*-value ≤ 0.05, Fishers Exact Test) denoted by asterisks

Table 12 summarizes the top 7 signatures with maximum relative risk (RR) of the minimum set of combination of gene expression patterns in predicting CAA status in men based on the NRF1 causal Bayesian network. The likelihood of CAA is almost 100% in a patient with the expression pattern of (low) NRF1 combined with GATC (low or no change), GDF9 (high or no change), SLC9B1 (high or no change), TOMM6 (high or no change), and ZNF135 (high or no change). However, a subject that has (high) NRF1 expression combined with GATC (high or no change), GDF9 (low or no change), SLC9B1 (low or no change), TOMM6 (low or no change), and ZNF135 (no change) has almost zero percent probability of CAA. These findings confirm the association of low NRF1 expression combined with the expression of its target genes results in the increased risk of developing CAA in men. The next step in our analysis was to determine the sequential order of how these genes are interacting leading to eventual disease outcome of CAA.

Gene Ordering Analysis by MCMC for Identifying Causal ID3 Signature Genes

To determine whether ID3 target genes are causally associated with CAA, we performed gene ordering analysis of Markov blanket genes and clinical variables of CAA using a MCMC method. For a particular MCMC gene order and structure, the *k* gene(s) that precede the disease state/survival (node) are causal for that node [30–32]. The weight of probable causal relationships is indicated by the thickness of the arc, and numbers in parentheses indicating reverse causal relationships. The ID3 causal Bayesian network for women is represented in an MCMC order in Fig. 16. The weight of connections (arcs) between nodes is determined by the thickness of the arc, and numbers in parentheses specify the reverse causal relationships. In temporal cortex samples derived from women patients with CAA (*n* = 72), all gene orders for ID3 causal Bayesian networks are observed in Table 13. In women CAA

Table 5 NRF1 and ID3 pathway activity in female CAA cases vs controls at Cerebellum

	Pathway activity	Total target gene no	No. significant genes	Coefficient	Odds ratio	P-Value	Total activity
NRF1	NRF1 Total Activity	10,246	2191	−0.00239	0.985	0.0165	Down*
	NRF1 Endothelial Signature	2272	728	−0.00203	0.9875	0.06	Down
	VEGF Signaling	56	19	−0.00465	0.972	0.583	Down
	Apoptosis	142	42	−0.00206	0.98726	0.828	Down
	Cell Cycle	76	18	0.007913	1.050406	0.519	Up
	Cell Senescence	67	23	−0.00871	0.9472	0.454	Down
	Epigenetic Chromatin Remodeling Factors	67	19	−0.006	0.9623	0.408	Down
	WNT Signaling Pathway	82	28	−0.0298	0.8310	0.0259	Down*
	Endothelial Cell Death	13	7	−0.04045	0.7778	0.0852	Down
	Endothelial Cell Growth and Proliferation	13	6	−0.03137	0.8229	0.131	Down
	Endothelial Cell Angiogenesis & Vasculogenesis	39	13	−0.000508	0.996	0.974	Down
	Endothelial Cell Cycle	28	9	−2.15 E -14	0.99	0.99	Down
	Endothelial Cell Development	13	3	0.0542	1.40	0.1035	Up
	Endothelial Cell Movement	21	3	0.0468	1.3379	0.2223	Up
ID3	ID3 Total Activity	9284	2114	0.001613	1.010074	0.237	Up
	ID3 Endothelial Signature	2619	890	−0.00162	0.990	0.125	Down
	VEGF Signaling	53	16	−0.00555	0.966	0.511	Down
	Apoptosis	211	53	0.00361	1.023	0.621	Up
	Cell Cycle	42	11	0.0077	1.049	0.438	Up
	Cell Senescence	49	12	0.00367	1.023	0.722	Up
	Epigenetic Chromatin Remodeling Factors	56	13	0.00618	1.039	0.411	Up
	WNT Signaling Pathway	69	23	−0.03102	0.825	0.022	Down*
	Endothelial Cell Death	10	4	−0.00423	0.974	0.842	Down
	Endothelial Cell Growth & Proliferation	14	9	−0.04977	0.733	0.040	Down*
	Endothelial Angiogenesis & Vasculogenesis	45	13	0.02748	1.186	0.2073	Up
	Endothelial Cell Cycle	19	6	0.0149	1.097	0.494	Down
	Endothelial Cell Development	14	3	0.0228	1.15	0.386	Up
	Endothelial Cell Movement	17	3	0.0291	1.1983	0.335	Up

Significant pathways (p -value ≤ 0.05 , Fishers Exact Test) denoted by asterisks

patients, there were 53 orders in total, with causal chain sub-structures for CAA consisting of ID3 (5.7% of all orders), DNAJC25 (16.9% of all orders), STAB2 (3.8% of all orders), APOE (5.7% of all orders), and MUC20 (13.2% of all orders) found across all orders. Furthermore, for clinical variables, causal sub-structures consisted of severity (3.7% of all orders) and Braak stage (3.7% of all orders). The consensus structure with the top gene order for the ID3-targeted CAA female gene network is shown in Fig. 16. In women CAA patients, fifty-three gene order sequences of ID3-regulatable genes predicted >99% of CAA disease outcomes (Table 13). In Fig. 16, the causal chain sub-structures connecting ID3 target genes, CAA, and clinical variables was APOE → BRAAK → MUC20 → CAA → ID3 → ID1. The connection between CAA → ID3 was at an 88.77% probability for causality and the probability of 5.99% causality when the causal order was reversed to ID3 → CAA. These results suggest that ID3 and its target genes are involved in the causality of CAA in women.

There is >99% probability that the direction of causality of CAA to order variables is shown in the optimal structure generated by MCMC order search between the following gene connections: CAA → TLR9, CAA → CELA2B, CAA → UTS2, CAA → ZNF135, CAA → MFSD2B, and CAA → SLC5A2. For clinical variables of interest, there was a >90% probability of the causal order between CAA → BRAAK, CAA → THAL, CAA → Severity, and CAA → THAL → APOE4. Furthermore, there was >90% probability of causality for gene connections involving BRAAK for each of the following orders: BRAAK → ATG10, BRAAK → IL9, BRAAK → U2AF1, and BRAAK → LYG2. A causal chain sub-structure involving NRF1 also occurred as CAA → BRAAK → DNAJC25 → NRF1 → ZNF135 was discovered. Amongst the 19 genes in the ID3 MCMC order search structure, several other causal connections were discovered, with several connections garnering >99% probability of causal directionality

Table 6 NRF1 and ID3 pathway activity in male CAA cases vs controls at Cerebellum

	Pathway activity	Total target gene no	No. significant genes	Coefficient	Odds ratio	P-Value	Total activity
NRF1	NRF1 Total Activity	10,246	4050	−0.005861	0.964	1.670 E ^{−20}	Down*
	NRF1 Endothelial Signature	2272	939	−0.0013	0.991	0.251	Down
	VEGF Signaling	56	19	0.004365	1.0275	0.579	Up
	Apoptosis	142	39	0.00619	1.0039	0.906	Up
	Cell Cycle	76	32	−0.015	0.908	0.130	Down
	Cell Senescence	67	17	0.0007	1.0042	0.934	Up
	Epigenetic Chromatin Remodeling Factors	56	19	0.0002	1.0012	0.981	Up
	WNT Signaling Pathway	82	27	−0.005	0.968	0.624	Down
	Endothelial Cell Death	13	5	−0.006	0.962	0.596	Down
	Endothelial Cell Growth and Proliferation	13	5	−0.00346	0.978	0.7577	Down
	Endothelial Cell Angiogenesis & Vasculogenesis	39	11	0.005116	1.032	0.794	Up
	Endothelial Cell Cycle	28	6	−9.37 E ^{−12}	1	1	Down
	Endothelial Cell Development	13	5	−0.0275	0.843	0.629	Down
	Endothelial Cell Movement	17	4	0.02511	1.1689	0.416	Up
ID3	ID3 Total Activity	9284	2191	0.00367	1.023	0.0106	Up*
	ID3 Endothelial Signature	2700	1180	−0.00617	0.962	9.13 E ^{−8}	Down*
	VEGF Signaling	53	16	0.00269	1.0169	0.725	Up
	Apoptosis	211	59	0.00963	1.0617	0.238	Up
	Cell Cycle	42	18	−0.00597	0.9636	0.500	Down
	Cell Senescence	49	12	0.000726	1.0045	0.923	Up
	Epigenetic Chromatin Remodeling Factors	67	21	−0.000698	0.995	0.941	Down
	WNT Signaling Pathway	69	25	−0.01809	0.893	0.0904	Down
	Endothelial Cell Death	10	6	0.0105	1.067	0.377	Up
	Endothelial Cell Growth & Proliferation	14	4	−0.0001	0.999	0.989	Down
	Endothelial Angiogenesis & Vasculogenesis	45	15	0.00512	1.23	0.134	Up
	Endothelial Cell Cycle	19	6	−0.0402	0.779	0.171	Down
	Endothelial Cell Development	14	1	0.1546	2.613	0.049	Up*
	Endothelial Cell Movement	21	5	0.0886	1.735	0.063	Up

Significant pathways (p -value ≤ 0.05 , Fishers Exact Test) denoted by asterisks

as generated through the machine search. In summary, the majority of MCMC gene orders found in Table 13 show that ID3 drives severity of CAA (ID3 $\rightarrow$ APOE4 $\rightarrow$ severity) in women.

To determine if ID3 signature genes are causal for CAA disease status in men, we performed causal ordering of Markov blanket genes as described previously. The consensus structure with the top MCMC ordered ID3 network for men with CAA is shown in Fig. 17. The weight of connections (arcs) between nodes is determined by the thickness of the arc, and numbers in parentheses specify the reverse causal probabilities. In temporal cortex samples derived from men with CAA ($n=80$), all gene orders for ID3 causal Bayesian networks are observed in Table 14. There were 49 orders in total, with these genes being the parent of CAA in the following amount of causal chain sub-structures percentages consisting of the following: ID3 (4.1% of all orders), APOE (48.9% of all orders), NEU3 (77.6% of all orders), USP50 (4.1% of all orders), SLC4A8 (4.1% of all orders), and KLC3 (6.1%

of all orders). In forty-nine gene order sequences, ID3 targeted genes predicted >99% of CAA disease outcomes (Table 14). There were three causal chain sub-structures connecting ID3 with CAA represented by: 1) NEU3 $\rightarrow$ CAA $\rightarrow$ SLC4A8 $\rightarrow$ ID3 $\rightarrow$ ID1, 2) NEU3 $\rightarrow$ CAA $\rightarrow$ BRAAK $\leftarrow$ ID3 $\rightarrow$ ID1, and 3) NEU3 $\rightarrow$ CAA $\rightarrow$ Thal $\rightarrow$ BRAAK $\leftarrow$ ID3 $\rightarrow$ ID1. The connection between BRAAK and ID3 is reversed meaning that causality for BRAAK and CAA is determined by ID3 at 8.37% causality. There is >90% probability that the direction of causality of CAA to order variables is shown in the optimal structure generated by MCMC order search between the following gene connections: CAA $\rightarrow$ NR1I3, CAA $\rightarrow$ CELA2B, CAA $\rightarrow$ UTS2, CAA $\rightarrow$ EFCAB10, CAA $\rightarrow$ PCDHGA1, CAA $\rightarrow$ SLC4A8, CAA $\rightarrow$ USP50, CAA $\rightarrow$ C16orf92, CAA $\rightarrow$ CBS, CAA $\rightarrow$ ZNF705A, CAA $\rightarrow$ FOXD4L4, and CAA $\rightarrow$ EPS8L3. A causal connection occurred between CAA $\rightarrow$ SLC4A8 (89.72%) and in reverse causality SLC4A8 $\rightarrow$ CAA (6.11%). For clinical variables of interest, there was a >99% probability for causality for CAA $\rightarrow$ THAL

Fig. 10 Bayesian network of ID3 target genes relevant to CAA. The 311 nodes in yellow represent signature genes that were significant ($p\text{-value} \leq 0.05, \pm 2\text{-Fold Change}$) in Benjamin Hochberg False Discovery Rate T-test. The five nodes in green indicate clinical and pathological information connected or related to the CAA nodes. The node in red represents CAA and the node in blue represents ID3

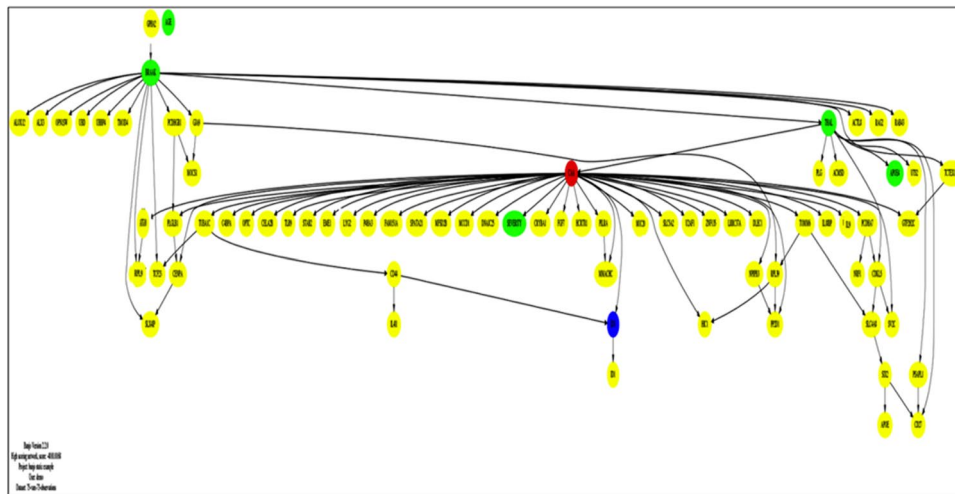
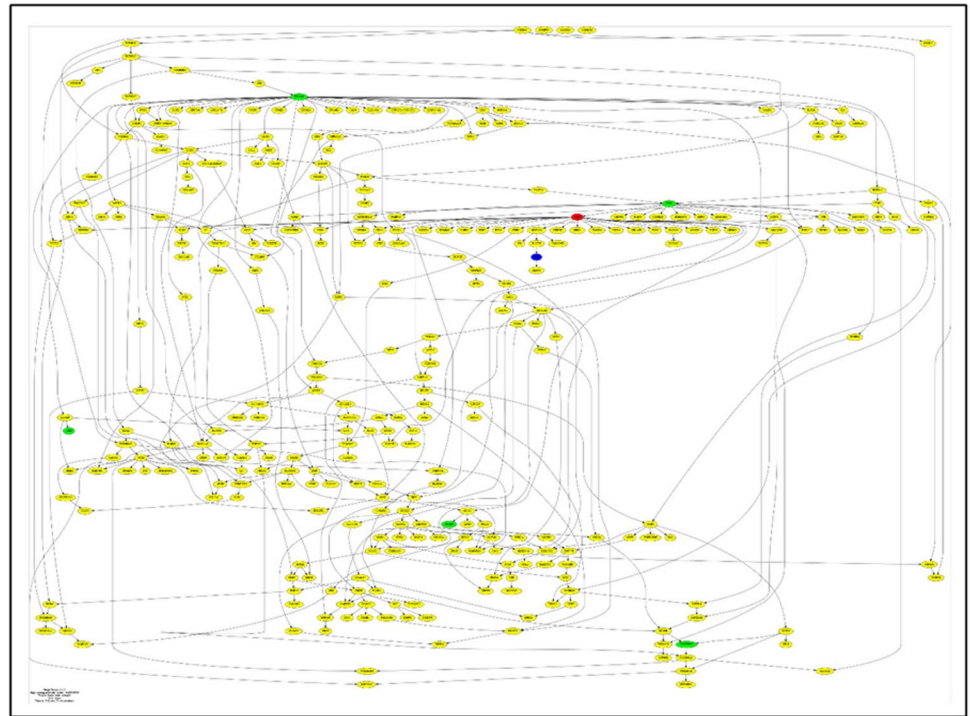


Fig. 11 Validation structure of Bayesian network ID3 target genes relevant to CAA in women. The variables in this structure were determined by a comparison of Markov blanket variables in the ID3 representative network (Fig. 10). The nodes in yellow represent 67 signature genes that were significant ($p\text{-value} \leq 0.05, \pm 2\text{-Fold Change}$)

in Benjamin Hochberg False Discovery Rate T-test. The five nodes in green indicate clinical and pathological information connected or related to the CAA nodes. The node in red represents CAA and the node in blue represents ID3, a direct child of CAA

and $CAA \rightarrow \text{Severity}$. For other clinical variables, probability for causality was as follows $CAA \rightarrow APOE4$ (35.91%) confirming findings from previous studies on the cohort that $APOE4$ allele plays a role in CAA in male patients (33) and $CAA \rightarrow BRAAK$ (14.39%). Furthermore, probability of causality for gene connections involving BRAAK for each of the following orders was as follows: $BRAAK \rightarrow SLC4A8$ (9.68%), $BRAAK \rightarrow APOE4$

(62.7%), $BRAAK \rightarrow ZNF705A$ (6.86%), $BRAAK \rightarrow \text{Severity}$ (1.79%), $BRAAK \rightarrow PCDHGA1$ (14.08%), $BRAAK \rightarrow C16orf92$ (0.97%). Several causal chain sub-structures involving NRF1 were discovered as: 1) $NEU3 \rightarrow CAA \rightarrow APOE4 \rightarrow NRF1$, 2) $NEU3 \rightarrow CAA \rightarrow BRAAK \rightarrow NRF1$, 3) $NEU3 \rightarrow CAA \rightarrow EFCAB10 \rightarrow NRF1$, and 4) $NEU3 \rightarrow CAA \rightarrow NRF1$ were all discovered. A multitude of

Table 7 ID3 Targeted Causal Markov blanket discovery genes of CAA in the structure with the best log-likelihood score

Parents	Children	Co- parents
Thal	ID3	MUC20
	Severity	PCDHGB1
	MUC20	PLGB1
	TUBA1C	PCDHA7
	STAB2	TOMM6
	MFSD2B	RPL39
	DNAJC25	CDKL5
	HIC1	TCTEX1D4
	LYG2	
	IL9	
	ATG10	
	RPL9	
	PLGB1	
	C4BPA	
	OPTC	
	CELA2B	
	TLR9	
	EME1	
	P4HA3	
	FAM151A	
	CENPA	
	SPATA21	
	MCCD1	
	CRYBA1	
	FGF7	
	HCRTR1	
	PILRA	
	MMACHC	
	SLC5A2	
	U2AF1	
	ZNF135	
	LRRC37A	
	DLEC1	
	RPL39	
	PP2D1	
	TOMM6	
	PCDHA7	
	CDKL5	
	SV2C	
	GTF2H2C	
	NPIPB3	

Variables in blue represent variables with a significant influence score (≥ 0.70) based on the BANJO scoring algorithm. These variables were inputted into the MCMC ordering algorithm

other causal gene connections ($n = 17$) were found in the optimal MCMC order search structure for the ID3 targeted CAA male network, with some connections garnering $> 99\%$ probability of causal directionality as generated through the machine search. The

majority of MCMC gene orders found in Table 14 show severity and CAA ordered before ID3. These findings provide further support that ID3 is a stronger driver of CAA severity in women.

Gene ordering Analysis by MCMC for Identifying Causal NRF1 Signature Genes

To determine whether NRF1 target genes are causally associated with CAA in men and women, we performed gene ordering of sex-specific CAA Markov blanket genes by the MCMC method. The causal Bayesian networks were generated using CAA temporal cortex samples from women ($n = 72$). The consensus gene order structure for the NRF1 causal Bayesian networks for women with CAA is shown in Fig. 18. Gene orders for NRF1 gene signatures are found in Table 15. In the NRF1 targeted network for women CAA patients, there were 57 signature gene orders in total, with these genes being the parent of CAA in the following amount of causal chain sub-structures percentages consisting of the following: ID3 (7.01% of all orders), CCDC103 (22.8% of all orders), SNORA70 (24.6% of all orders), CFLARS1 (10.5% of all orders), and RNU6ATAC (5.3% of all orders). Clinical variables showed causal sub-structures with BRAAK (100% of all orders) and THAL stage (1.8% of all orders). Fifty-seven gene order sequences of NRF1-regulatable genes predicted $> 99\%$ of CAA disease outcomes in women (Table 15). There were three causal chain sub-structures connecting NRF1 with CAA represented by: 1) CAA $\rightarrow$ IL7 $\rightarrow$ NRF1, 2) CAA $\rightarrow$ SNORA81 $\rightarrow$ NRF1, and 3) CAA $\rightarrow$ PTC1 $\rightarrow$ SNORA81 $\rightarrow$ NRF1. There is $> 90\%$ probability that the direction of causality of CAA to order variables is shown in the optimal structure (Fig. 18) generated by MCMC order search between the following gene connections: CAA $\rightarrow$ ID3, CAA $\rightarrow$ RAET1G, CAA $\rightarrow$ SNORA52, CAA $\rightarrow$ RUSC1A1, CAA $\rightarrow$ MMACHC, CAA $\rightarrow$ TOMM6, CAA $\rightarrow$ ZNF670ZNF695, and CAA $\rightarrow$ IL7. Causal gene connections also occurred between CAA $\rightarrow$ ID1 (8.16%), CAA $\rightarrow$ PTC1 (61.80% causality), CAA $\rightarrow$ LONL3 (6.04% causality), CAA $\rightarrow$ STAG3 (4.96% causality), CAA $\rightarrow$ CFLARAS1 (3.8% causality), CAA $\rightarrow$ MCM3A-PAS1 (4.91% causality), CAA $\rightarrow$ RNU6ATAC (3.56% causality), and CAA $\rightarrow$ SNORA81 (34.52% causality). For clinical variables of interest, there was a $> 90\%$ probability for causality for CAA $\rightarrow$ THAL, CAA $\rightarrow$ Severity, and BRAAK $\rightarrow$ CAA meaning that BRAAK stage played a significant role in CAA status in women with CAA. Furthermore, APOE4 allele status contributed to causality of CAA outcome as we observed CAA $\rightarrow$ APOE4 (6.98% causality) according to the optimal model. Our findings showed that BRAAK stage played a crucial role in the NRF1 targeted networks. Probability of causality for gene connections involving BRAAK for each of the following orders was as follows: BRAAK $\rightarrow$ LOXL3 (95.36%), BRAAK $\rightarrow$ STAG3 (89.67%),

Fig. 12 Causal Bayesian network of NRF1 target genes relevant to CAA. The 283 nodes in yellow represent signature genes that were significant ($p\text{-value} \leq 0.05, \pm 2\text{-Fold Change}$) in Benjamin Hochberg False Discovery Rate t-test. The five nodes in green indicate clinical and pathological information connected or related to the CAA nodes. The node in red represents CAA and the node in blue represents NRF1

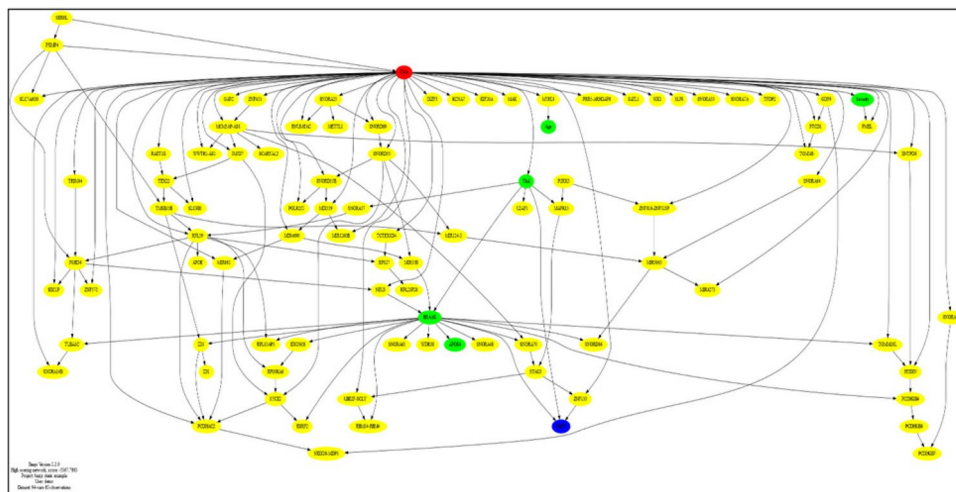
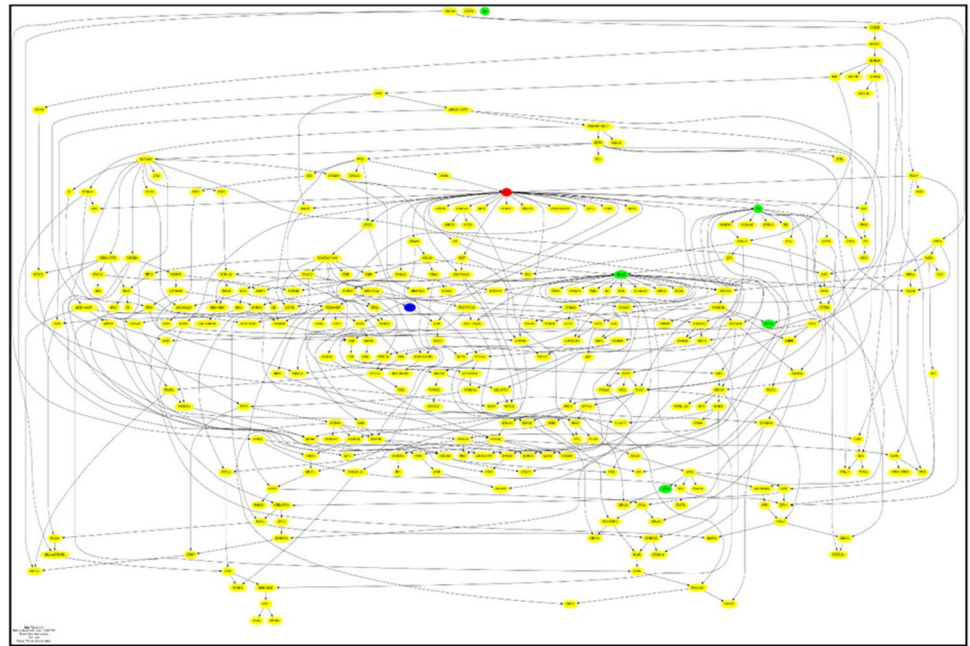


Fig. 13 Validation structure of Bayesian network NRF1 target genes relevant to CAA. The variables in this structure were determined by a comparison of Markov blanket variables in the original NRF1 representative network (Fig. 12). The nodes in yellow represent 87 genes that were significant ($p\text{-value} \leq 0.05, \pm 2\text{-Fold Change}$) in Benjamin

Hochberg False Discovery Rate T-test. The five nodes in green indicate clinical and pathological information connected or related to the CAA nodes. The node in red represents CAA and the node in blue represents ID3, a direct child of CAA

BRAAK $\rightarrow$ MCM3APAS1 (96.44%), BRAAK $\rightarrow$ MCMDC2 (95.73%), BRAAK $\rightarrow$ MMACHC (2.08%), BRAAK $\rightarrow$ SNORA81 (66.93%), BRAAK $\rightarrow$ CCDC103 (92.27%), BRAAK $\rightarrow$ ZNF670ZNF695 (2.25%), BRAAK $\rightarrow$ SNORA70 (96.9%), BRAAK $\rightarrow$ RAET1G (1.95%), BRAAK $\rightarrow$ CFLARAS1 (97.35% causality), BRAAK $\rightarrow$ SNORA52 (5.85% causality), BRAAK $\rightarrow$ RNU6ATAC (96.44% causality), and BRAAK $\rightarrow$ FANCA (96.44% causality). Furthermore, BRAAK was connected to clinical variables such as Thal (92.04%) and Severity

(3.39%). Several causal chain sub-structures involving connecting CAA causality with NRF1 were discovered as: 1) CAA $\rightarrow$ IL7 $\rightarrow$ NRF1, 2) CAA $\rightarrow$ SNORA81 $\rightarrow$ NRF1, 3) CAA $\rightarrow$ PTC1 $\rightarrow$ SNORA81 $\rightarrow$ NRF1, and 4) CAA $\rightarrow$ LOXL3 $\rightarrow$ NRF1 were all discovered because of the ordering search. Various other causal gene connections ($n=22$ genes) optimal were found amongst the MCMC order search structure for the NRF1 targeted CAA female network, with several connections garnering $>99\%$ (Fig. 18) probability of causal directionality as generated through the

Table 8 NRF1 targeted causal Markov Blanket discovery nodes of CAA in the structure with the best log-likelihood score

Parents	Children	Co- parents
SERHL	Thal	ID3
PXMP4	Severity	GDF9
	ZNF570	Severity
	RPL39	MCM3AP-AS1
	SNORA31	RPL39
	GATC	PTCD1
	SLC9B1	ENTPD8
	SNORA64	P2RX5
	TRIM44	STAG3
	ZNF135	BRAAK
	SATL1	MIR3665
	GDF9	MIR339
	MCM3AP-AS1	SNORD35B
	TOMM6	SNORD89
	SNORA33	PGBD4
	EXOSC6	MIR661
	SNORA25	TEX22
	RNU6ATAC	SNORD55
	SNORD19	ZNF431
	KCNA7	SNORA57
	ZNF491	TMSB15B
	RAET1G	MIR661
	SNORA148	SYCE2
	SEC1P	TUBA1C
	MAK	PXMP4
	KIF30A	SNORA25
	MIR661	
	PCDHAC2	
	MYH11	
	PRR5-ARHGAP8	
	SNORA7A	
	SIK1	
	SLPI	
	PTCD1	
	PMEL	
	ENTPD8	
	SNORA81	
	IKZF2	
	HYDIN	
	TOMM20L	
	MIR4273	
	ZNF816-ZNF321P	
	MIR339	
	SNORD35B	
	SNORD89	
	POLR2J2	
	MIR1260B	
	MIR4690	
	WWTR1-AS1	

Table 8 (continued)

Parents	Children	Co- parents
	JMJD7	
	UBE2F-SCLY	
	TCTEX1D4	
	SLC7A6OS	

Variables in blue represent variables with a significant influence score (≥ 0.70) based on the BANJO scoring algorithm. These variables were inputted into the MCMC ordering algorithm

Table 9 Gene combination of CAA lifetime risk scores for CAA with highest probabilities for CAA in ID3 female network

Lifetime CAA risk in women from ID3 network		
Single genes and gene combinations	P(CAA)	Fold change
Baseline risk with no gene level evidence set	0.08	1
ID3 ↑	0.483	6.035
STAB2 ↑	0.776	9.694
MUC20 ↑	0.789	9.861
TUBA1C ↑	0.801	10.008
IL9 ↑	0.811	10.139
HIC1 ↓	0.811	10.14
ZNF135 ↑	0.812	10.15
MFSD2B ↑	0.821	10.257
LYG2 ↑	0.827	10.267
CELA2B ↑	0.829	10.366
TLR9 ↑	0.837	10.463
DNAJC25 ↑	0.844	10.552
U2AF1 ↑	0.844	10.552
ATG10 ↑	0.851	10.634
SLC5A2 ↑	0.862	10.776
UTS2 ↑	0.867	10.842
ID3 ↑, STAB2 ↑	0.974	12.18
ID3 ↑, MUC20 ↑	0.976	12.203
ID3 ↑, TUBA1C ↑	0.978	12.224
ID3 ↑, LYG2 ↑	0.979	12.238
ID3 ↑, HIC1 ↓	0.979	12.24
ID3 ↑, ZNF135 ↑	0.979	12.241
ID3 ↑, IL9 ↑	0.979	12.241
ID3 ↑, MFSD2B ↑	0.981	12.257
ID3 ↑, CELA2B ↑	0.981	12.268
ID3 ↑, DNAJC25 ↑	0.982	12.271
ID3 ↑, TLR9 ↑	0.982	12.28
ID3 ↑, U2AF1 ↑	0.984	12.294
ID3 ↑, ATG10 ↑	0.984	12.301
ID3 ↑, SLC5A2 ↑	0.986	12.321
ID3 ↑, UTS2 ↑	0.986	12.329

Fig. 14 Causal Bayesian network Structure Using GeNIe. The ID3 target representative network of CAA, with 1st MB genes (yellow) and clinical variables (orange). States 0 and 1 represent the probability of CAA, APOE4, and BRAAK stage. States 0, 1, 2 in Thal and severity represent low, moderate, and severe probability for the clinical node. In gene nodes states 0, 1, and 2 represent the gene expression levels of low, no change, and high, respectively. **A** Marginal probability for developing CAA (red) goes up to 91% when ID3 is modified to a state of overexpression. (ID3 100% Upregulated). **B** Marginal probability for developing CAA goes down to 8% when ID3 is modified to downregulation. **C** Marginal probability for developing CAA remains at 49% when ID3 is at normal expression

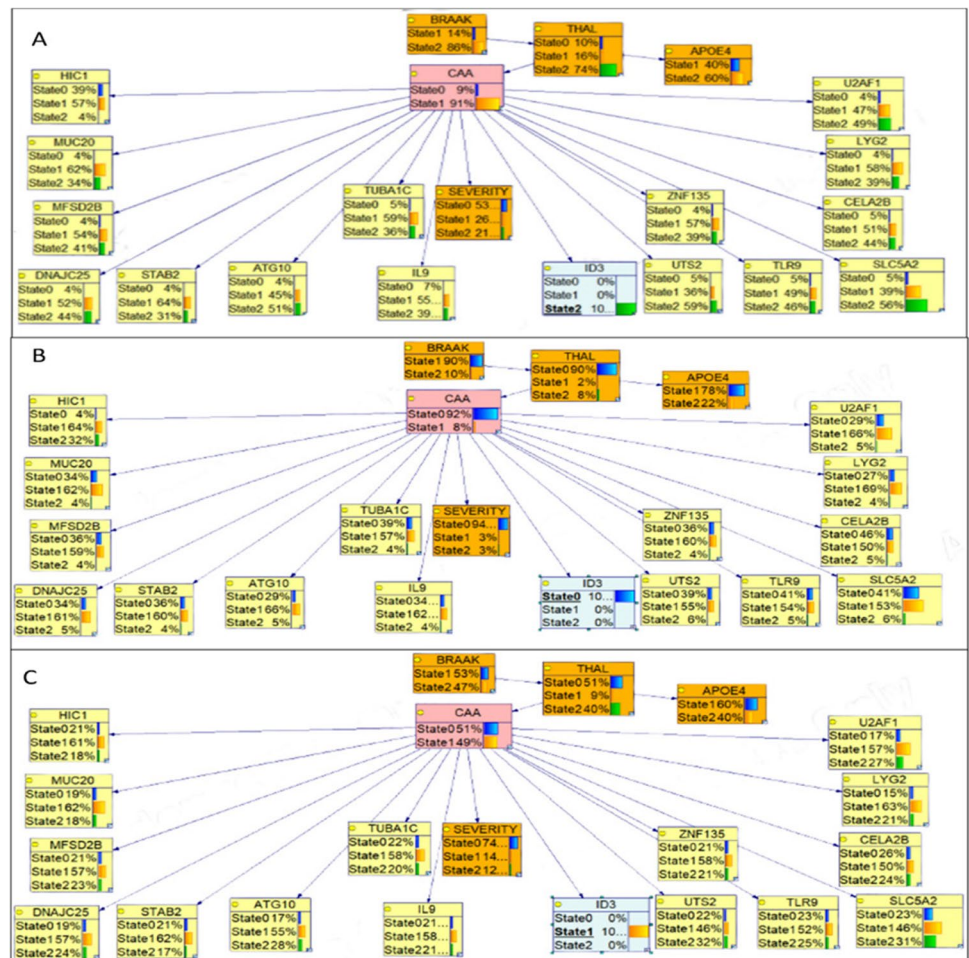


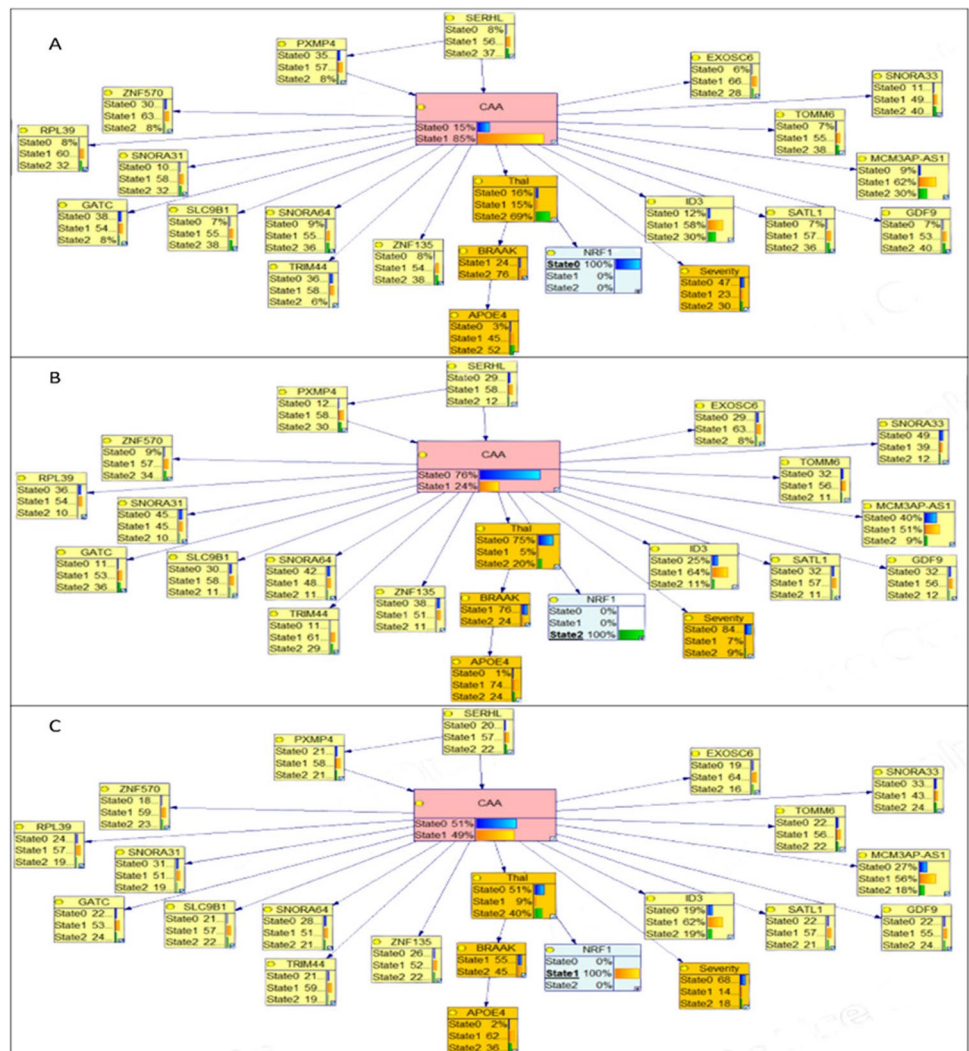
Table 10 Summary of the top 7 maximum relative risk (RR) ratios of the minimum set of combination of gene expression patterns for the ID3 representative network

Gene expression patterns	P (CAA = 1 Modified Expression Levels)	RR
ATG10 [2], ID3 [2], SLC5A2 [2], U2AF1 [2], UTS2 [2]	0.999999999	1.36E + 06
ATG10 [1], ID3 [0], SLC5A2 [0], U2AF1 [0], UTS2 [0]	7.32956E-07	
ATG10 [2], ID3 [2], SLC5A2 [1], U2AF1 [2], UTS2 [2]	0.999999999	3.94E + 04
ATG10 [0], ID3 [0], SLC5A2 [1], U2AF1 [1], UTS2 [0]	0.000025374	
ATG10 [2], ID3 [2], SLC5A2 [2], U2AF1 [1], UTS2 [2]	0.999999999	3.34E + 04
ATG10 [0], ID3 [0], SLC5A2 [1], U2AF1 [0], UTS2 [1]	2.99185E-05	
ATG10 [2], ID3 [2], SLC5A2 [2], U2AF1 [2], UTS2 [1]	0.999999999	2.20E + 03
ATG10 [1], ID3 [0], SLC5A2 [0], U2AF1 [1], UTS2 [1]	0.000454961	
ATG10 [2], ID3 [2], SLC5A2 [1], U2AF1 [1], UTS2 [2]	0.999956000	1.83E + 03
ATG10 [1], ID3 [0], SLC5A2 [1], U2AF1 [1], UTS2 [0]	0.000546380	
ATG10 [1], ID3 [2], SLC5A2 [1], U2AF1 [2], UTS2 [2]	0.999950000	1.55E + 03
ATG10 [1], ID3 [0], SLC5A2 [1], U2AF1 [0], UTS2 [1]	0.000644176	
ATG10 [1], ID3 [2], SLC5A2 [2], U2AF1 [1], UTS2 [1]	0.994546000	6.82E + 01
ATG10 [1], ID3 [0], SLC5A2 [1], U2AF1 [1], UTS2 [1]	0.014577300	

[0] = low expression, [1] = no change in the expression, [2] = high expression

Predictive probability of CAA (1 = CAA, 0 = control) was calculated following modification of Markov blanket genes using the PREDICT algorithm

Fig. 15 NRF1 Causal Bayesian network Structure using GeNIe Modeler. The NRF1 targeted representative network of CAA, with 1st MB genes (yellow) and clinical variables (orange). States 0 and 1 represent the probability of CAA, APOE4, and BRAAK stage. States 0, 1, 2 in Thal and severity represent low, moderate, and severe probability for the clinical node. In gene nodes states 0, 1, and 2 represent the gene expression levels of low, no change, and high, respectively. **A** Marginal probability for developing CAA (red) goes up to 85% when NRF1 is modified to a state of downregulation. (NRF1 100% downregulated). **B** Marginal probability for developing CAA goes down to 24% when NRF1 is modified to downregulation. **C** Marginal probability for developing CAA remains at 49% when NRF1 is at normal expression



machine search. All causal order discoveries are detailed in Table 15. In summary, the consensus MCMC NRF1 gene ordered structure, there was no direct arc connection between CAA and NRF1 in women.

Next we evaluated the MCMC ordered causal NRF1 gene signatures in men with CAA. The consensus structure with the top gene order for the NRF1-targeted gene network in men is illustrated in Fig. 19. In men with CAA, twenty-four different gene order sequences of NRF1-regulatable genes predicted > 99% of CAA outcomes with NRF1 ranked as the parent (causal node) of CAA in all order outcomes (Table 16). In temporal cortex samples derived from men CAA patients (n = 80), all gene orders for NRF1 causal Bayesian networks are observed in Table 16. There were 24 gene signature orders in total, with these genes being the causal parent of CAA in the following amount of causal chain sub-structures percentages consisting of the following: NRF1 (100% of all orders), MCM3APS1 (100% of all orders), ID1 (58.3% of all orders). The NRF1-male targeted

network was unique in that NRF1 is the causal parent of CAA as compared to the ID3 male/female and NRF1 female CAA regulated networks.

Furthermore, in terms of Markov blanket gene connections the strongest probability for causality was found in the NRF1 network for men. The causal parents of CAA showed significant causality of > 99% probability in NRF1 → CAA and MCM3APAS1 → CAA connections. There was > 90% probability that the direction of causality of CAA to order variables is shown in the optimal structure generated by MCMC order search between the following: CAA → GATC, CAA → RNU6ATAC, CAA → SNORA64, CAA → RPL39, CAA → SNORA31, CAA → SATL1, CAA → SLC9B1, CAA → SNORD65, CAA → ZNF570, CAA → GDF9, CAA → SNORA33, and CAA → TOMM6. Further causal gene connections occurred between CAA → EXOSC6 (16.13%), CAA → ID1 (0.30%), CAA → ID3 (0.58%), and CAA → ZNF135 (2.49%). In terms of causal connections between CAA and clinical variables, several

Table 11 Gene combination CAA lifetime risk scores for CAA in NRF1 male network

Lifetime CAA risk in men from NRF1 network		
Single genes and gene combinations	P(CAA)	Fold change
Baseline risk with no gene level evidence set	0.08	1
NRF1 ↓	0.322	4.024
ID3 ↑	0.481	6.014
PXMP4 ↓	0.573	7.166
SERHL ↑	0.586	7.326
EXOSC6 ↑	0.775	9.691
MCM3AP-AS1 ↑	0.788	9.845
ZNF570 ↓	0.789	9.857
RPL39 ↑	0.799	9.984
SNORA31 ↑	0.8	9.994
TRIM44 ↓	0.818	10.221
SATL1 ↑	0.818	10.225
GATC ↓	0.826	10.323
ZNF135 ↑	0.826	10.324
GDF9 ↑	0.826	10.324
TOMM6 ↑	0.829	10.359
SLC9B1 ↑	0.829	10.363
NRF1 ↓, ID3 ↑	0.836	10.451
SNORA33 ↑	0.836	10.452
SNORA64 ↑	0.837	10.463
NRF1 ↓, PXMP4 ↓	0.881	11.01
NRF1 ↓, SERHL ↑	0.886	11.078
NRF1 ↓, EXOSC6 ↑	0.95	11.874
NRF1 ↓, MCM3AP-AS1 ↑	0.953	11.916
NRF1 ↓, ZNF570 ↓	0.953	11.916
NRF1 ↓, RPL39 ↑	0.956	11.952
NRF1 ↓, SNORA33 ↑	0.956	11.952
NRF1 ↓, SNORA31 ↑	0.957	11.968
NRF1 ↓, SATL1 ↑	0.961	12.013
NRF1 ↓, SNORA64 ↑	0.963	12.032
NRF1 ↓, GATC ↓	0.963	12.039
NRF1 ↓, ZNF135 ↑	0.963	12.039
NRF1 ↓, SLC9B1 ↑	0.963	12.041
NRF1 ↓, TRIM44 ↓	0.965	12.068

connections were significant including: CAA → APOE4 (59.11% causality), CAA → Severity (99.39% causality), and CAA → BRAAK (31.36% causality). The strong causality between CAA and APOE4 in the male CAA network amongst NRF1 regulated genes confirms the significance of both NRF1 targeting of the APOE4 allele amongst male CAA cases, reported in previous studies examining the transcription factor and its targeting of an exon of APOE4 [17, 33].

Furthermore, probability of causality for gene connections involving BRAAK for each of the following orders was as follows: BRAAK → ZNF135 (97.09%), ZNF135 → BRAAK (0.52%), BRAAK → SNORA33

(96.50%), SNORA33 → BRAAK (1.15%), BRAAK → EXOSC6 (84.57%), EXOSC6 → BRAAK (0.33%), BRAAK → ID3 (98.88%), ID3 → BRAAK (1.12%), SLC9B1 → BRAAK (13.21%), SNORD55 → BRAAK (3.49%), ZNF570 → BRAAK (9.33%), RNU-6ATAC → BRAAK (1.83%), RPL39 → BRAAK (0.01%). Clinical connection between BRAAK stage and APOE4 status was found to be significant in men with CAA following causal order search as: BRAAK → APOE4 (2.92%) and reverse causality of APOE4 → BRAAK (0.43%) was attributed in the MCMC search. Several causal chain sub-structures involving ID3 were discovered in the NRF1 network with the most significant being the following: 1) NRF1 → CAA → BRAAK → ID3 → ID1, 2) NRF1 → CAA → All causal parents of BRAAK → BRAAK → ID3 → ID1, 3) MCM3APAS1 → CAA → BRAAK → ID3 → ID1, 4) MCM3APAS1 → CAA → All causal parents of BRAAK → BRAAK → ID3 → ID1 were all discovered. Various other causal connections were found in the optimal MCMC order search structure between the (n = 19) causal gene targets in the NRF1 regulated CAA male network, with several connections garnering > 99% probability of causal directionality as generated through the machine search as evidenced below in Fig. 19. In summary, MCMC gene ordering showed that NRF1 is a sex-specific driver of CAA in men while ID3 contributes to CAA severity in women.

Discussion

Cerebral amyloid angiopathy is a small vessel disease that causes intracerebral hemorrhages and vascular related cognitive decline. The exact molecular mechanism of this disease remains to be fully understood. Amyloid-β induced oxidative stress is one of the mechanisms being considered for cell and tissue injury in CAA [34]. Exposure to amyloid induces apoptosis in cerebrovascular smooth muscle and endothelial cells [35, 36]. Amyloid-β treatment has been shown to increase the ID family of proteins in cortical cells [14] but the contribution of ID3 to CAA are not known. Furthermore, APOE4 is a genetic risk factor for CAA associated with higher accumulation of amyloid-β in cerebral vessels. Severe CAA has a prevalence of 48% in AD patients [37] and women with APOE4 are at higher risk of AD. There is a DNA binding motif for NRF1 in the APOE4 gene, however, its contribution to CAA has not been investigated until now. Despite tremendous progress in understanding how amyloid-β and APOE4 contribute to CAA, gaps remain in understanding the molecular drivers of this small vessel disease. In this study, we have focused our efforts on uncovering ID3 and NRF1-driven transcriptional gene signatures of APOE4 carrying CAA patients. Our findings showed high

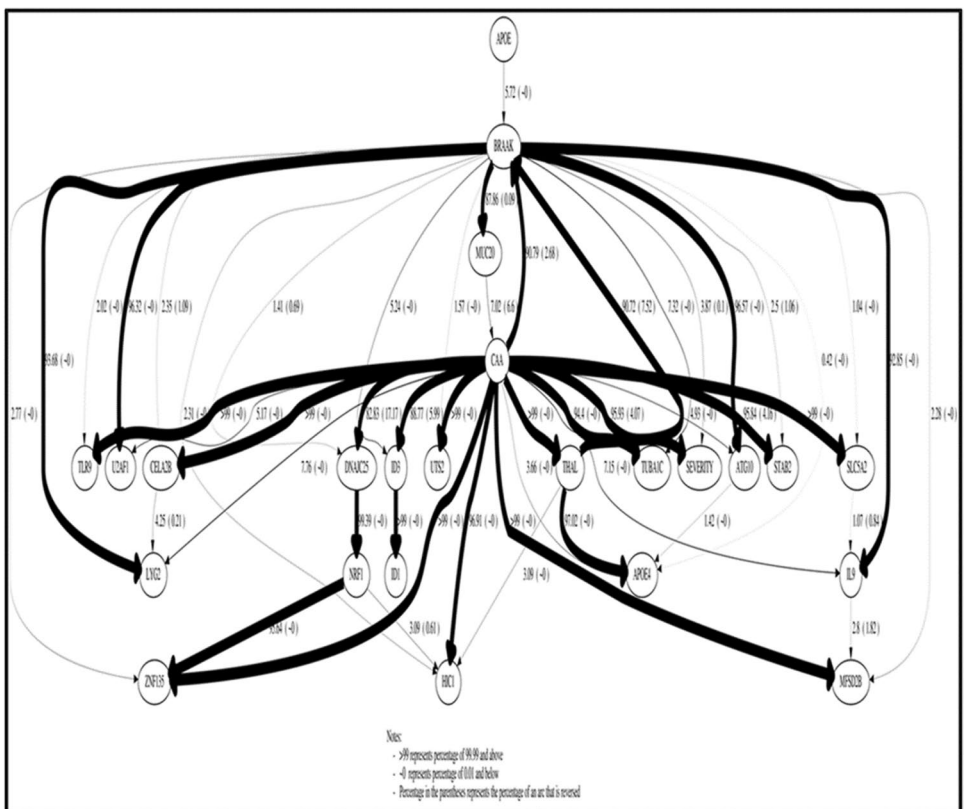
Table 12 Summary of the top 7 maximum relative risk (RR) ratios of the minimum set of combination of gene expression patterns for the NRF1 representative network

Gene expression patterns	P (CAA = 1 Modified Expression Levels)	RR
GATC [0], GDF9 [2], NRF1 [0], SLC9B1 [1], TOMM6 [2], ZNF135 [2]	0.999999999	8.20E + 07
GATC [2], GDF9 [0], NRF1 [2], SLC9B1 [0], TOMM6 [0], ZNF135 [1]	0.000000012	
GATC [0], GDF9 [2], NRF1 [0], SLC9B1 [2], TOMM6 [2], ZNF135 [2]	0.999999999	5.22E + 05
GATC [2], GDF9 [0], NRF1 [2], SLC9B1 [0], TOMM6 [1], ZNF135 [1]	0.000001917	
GATC [0], GDF9 [2], NRF1 [0], SLC9B1 [1], TOMM6 [2], ZNF135 [2]	0.999999999	1.21E + 04
GATC [2], GDF9 [1], NRF1 [2], SLC9B1 [1], TOMM6 [0], ZNF135 [1]	0.000082300	
GATC [1], GDF9 [2], NRF1 [0], SLC9B1 [2], TOMM6 [1], ZNF135 [2]	0.999999999	6.06E + 02
GATC [1], GDF9 [0], NRF1 [2], SLC9B1 [1], TOMM6 [1], ZNF135 [1]	0.001654240	
GATC [1], GDF9 [1], NRF1 [0], SLC9B1 [2], TOMM6 [1], ZNF135 [2]	0.999923000	2.45E + 02
GATC [2], GDF9 [1], NRF1 [2], SLC9B1 [1], TOMM6 [1], ZNF135 [1]	0.004081230	
GATC [1], GDF9 [1], NRF1 [0], SLC9B1 [2], TOMM6 [1], ZNF135 [1]	0.999692000	5.02E + 00
GATC [1], GDF9 [1], NRF1 [2], SLC9B1 [1], TOMM6 [1], ZNF135 [1]	0.199246000	
GATC [1], GDF9 [1], NRF1 [0], SLC9B1 [2], TOMM6 [1], ZNF135 [1]	0.996135000	5.00E + 00
GATC [1], GDF9 [1], NRF1 [2], SLC9B1 [1], TOMM6 [1], ZNF135 [1]	0.199246000	

[0]=low expression, [1]=no change in the expression, [2]=high expression

Predictive probability of CAA (1 = CAA, 0 = control) was calculated following modification of Markov blanket genes using the PREDICT algorithm

Fig. 16 MCMC ordered ID3 network for women with CAA



expression of ID3 mRNA coupled with low NRF1 mRNA levels in the temporal cortex of CAA patients. It is worth highlighting that low NRF1 mRNA expression was found in men with severe CAA. Also, men with severe CAA had

higher ID3 expression with a Braak stage of 5/6. Since men are more frequently diagnosed with severe CAA, these results suggest that ID3 and NRF1 may contribute to sex specific severity of CAA. In women, we showed higher ID3

Table 13 Markov blanket MCMC causal gene orders of CAA female cases/controls and ID3 targets

Order rank	Order probability	Cumulative order probability	1	2	3	4	5	6	7	8	9	10	11
1	0.02845	0.02845	CAA	CELA2B	THAL	BRAAK	TUBA1C	DNAJC25	NRF1	MUC20	ID3	ID1	TLR9
2	0.02763	0.05608	CAA	CELA2B	THAL	BRAAK	TUBA1C	DNAJC25	NRF1	MUC20	ID3	ID1	TLR9
3	0.02762	0.08370	CAA	CELA2B	THAL	BRAAK	TUBA1C	DNAJC25	NRF1	MUC20	ID3	ID1	TLR9
4	0.02760	0.11131	CAA	CELA2B	THAL	BRAAK	TUBA1C	DNAJC25	NRF1	MUC20	ID3	ID1	TLR9
5	0.02549	0.13679	CAA	CELA2B	THAL	DNAJC25	BRAAK	ATG10	U2AF1	MUC20	NRF1	TUBA1C	ID3
6	0.02543	0.16223	CAA	CELA2B	THAL	DNAJC25	BRAAK	ATG10	U2AF1	MUC20	ID3	TUBA1C	NRF1
7	0.02416	0.18639	CAA	CELA2B	THAL	DNAJC25	BRAAK	TLR9	U2AF1	APOE	NRF1	TUBA1C	ID3
8	0.02288	0.20927	CAA	CELA2B	THAL	DNAJC25	BRAAK	TLR9	U2AF1	APOE	NRF1	TUBA1C	ID3
9	0.02281	0.23208	CAA	CELA2B	THAL	DNAJC25	BRAAK	TLR9	U2AF1	APOE	NRF1	TUBA1C	ID3
10	0.02213	0.25421	CAA	SLC5A2	THAL	BRAAK	ID3	IL9	MFS2D2B	TUBA1C	CELA2B	DNAJC25	SEVERITY
11	0.02193	0.27613	CAA	CELA2B	THAL	DNAJC25	BRAAK	ATG10	U2AF1	APOE	NRF1	TUBA1C	ID3
12	0.02087	0.29700	DNAJC25	CAA	CELA2B	THAL	STAB2	BRAAK	ATG10	ID3	MUC20	U2AF1	TUBA1C
13	0.02079	0.31779	STAB2	CAA	CELA2B	THAL	DNAJC25	BRAAK	ATG10	U2AF1	MUC20	ID3	TUBA1C
14	0.02079	0.33858	DNAJC25	CAA	CELA2B	THAL	STAB2	BRAAK	ATG10	ID3	MUC20	U2AF1	TUBA1C
15	0.02077	0.35935	STAB2	CAA	CELA2B	THAL	DNAJC25	BRAAK	ATG10	ID3	MUC20	U2AF1	TUBA1C
16	0.02035	0.37970	SEVERITY	CAA	CELA2B	THAL	BRAAK	TUBA1C	DNAJC25	NRF1	MUC20	ID3	ID1
17	0.02034	0.40005	SEVERITY	CAA	CELA2B	THAL	BRAAK	TUBA1C	DNAJC25	NRF1	MUC20	ID3	ID1
18	0.02024	0.42029	DNAJC25	CAA	CELA2B	THAL	STAB2	BRAAK	ATG10	ID3	MUC20	U2AF1	TUBA1C
19	0.02014	0.44043	CAA	SLC5A2	THAL	BRAAK	CELA2B	SEVERITY	MFS2D2B	TUBA1C	ID3	DNAJC25	LYG2
20	0.02014	0.46057	DNAJC25	CAA	CELA2B	THAL	STAB2	BRAAK	ATG10	ID3	MUC20	U2AF1	TUBA1C
21	0.02010	0.48067	CAA	SLC5A2	THAL	BRAAK	ID3	SEVERITY	MFS2D2B	TUBA1C	CELA2B	DNAJC25	LYG2
22	0.02007	0.50074	CAA	SLC5A2	THAL	BRAAK	ID3	SEVERITY	MFS2D2B	TUBA1C	CELA2B	DNAJC25	LYG2
23	0.02002	0.52076	CAA	SLC5A2	THAL	BRAAK	CELA2B	STAB2	MFS2D2B	TUBA1C	ID3	DNAJC25	LYG2
24	0.02001	0.54077	ID3	APOE	CAA	HIC1	THAL	BRAAK	ATG10	TLR9	DNAJC25	UTS2	NRF1
25	0.01995	0.56071	ID3	APOE	CAA	HIC1	THAL	BRAAK	ATG10	CELA2B	DNAJC25	UTS2	NRF1
26	0.01994	0.58065	ID3	APOE	CAA	HIC1	THAL	BRAAK	ATG10	CELA2B	DNAJC25	UTS2	NRF1
27	0.01964	0.60029	CAA	SLC5A2	THAL	BRAAK	ID3	SEVERITY	MFS2D2B	TUBA1C	CELA2B	DNAJC25	LYG2
28	0.01952	0.61981	DNAJC25	CAA	CELA2B	THAL	STAB2	BRAAK	ATG10	ID3	MUC20	SLC5A2	TUBA1C
29	0.01902	0.63883	CAA	SLC5A2	THAL	BRAAK	ID3	IL9	HIC1	TUBA1C	CELA2B	DNAJC25	SEVERITY
30	0.01886	0.65769	DNAJC25	CAA	CELA2B	THAL	STAB2	BRAAK	ATG10	ID3	MUC20	SLC5A2	TUBA1C
31	0.01855	0.67624	CAA	SLC5A2	THAL	BRAAK	ID3	APOE	HIC1	CELA2B	LYG2	SEVERITY	IL9
32	0.01850	0.69474	CAA	SLC5A2	THAL	BRAAK	ID3	APOE	HIC1	CELA2B	LYG2	SEVERITY	IL9
33	0.01850	0.71324	CAA	SLC5A2	THAL	BRAAK	ID3	APOE	HIC1	CELA2B	LYG2	SEVERITY	DNAJC25
34	0.01838	0.73162	CAA	SLC5A2	THAL	BRAAK	ID3	APOE	HIC1	TUBA1C	CELA2B	DNAJC25	SEVERITY
35	0.01801	0.74963	DNAJC25	CAA	CELA2B	STAB2	THAL	BRAAK	ATG10	ID3	MUC20	SLC5A2	TUBA1C
36	0.01788	0.76750	CAA	SLC5A2	THAL	BRAAK	ID3	LYG2	MFS2D2B	TUBA1C	CELA2B	DNAJC25	SEVERITY

Table 13 (continued)

Order rank	Order prob-ability	Cumulative order probability	1	2	3	4	5	6	7	8	9	10	11
37	0.01757	0.78507	CAA	SLC5A2	THAL	BRAAK	ID3	IL9	HIC1	DNAJC25	LYG2	NRF1	UTS2
38	0.01741	0.80248	CAA	SLC5A2	THAL	BRAAK	ID3	IL9	HIC1	DNAJC25	LYG2	NRF1	UTS2
39	0.01740	0.81988	CAA	SLC5A2	THAL	BRAAK	ID3	IL9	HIC1	DNAJC25	LYG2	SEVERITY	UTS2
40	0.01691	0.83679	CAA	BRAAK	THAL	DNAJC25	CELA2B	TLR9	U2AF1	APOE	NRF1	TUBA1C	ID3
41	0.01691	0.85370	CAA	BRAAK	THAL	DNAJC25	CELA2B	TLR9	U2AF1	APOE	NRF1	TUBA1C	ID3
42	0.01674	0.87044	DNAJC25	CAA	CELA2B	STAB2	THAL	BRAAK	ATG10	ID3	UTS2	SLC5A2	TUBA1C
43	0.01653	0.88697	DNAJC25	CAA	CELA2B	STAB2	THAL	BRAAK	ATG10	ID3	UTS2	SLC5A2	TUBA1C
44	0.01602	0.90299	CAA	BRAAK	THAL	DNAJC25	CELA2B	TLR9	U2AF1	APOE	NRF1	TUBA1C	ID3
45	0.01341	0.91639	BRAAK	CAA	APOE	THAL	ID3	UTS2	CELA2B	SLC5A2	ATG10	LYG2	HIC1
46	0.01340	0.92980	BRAAK	CAA	APOE	THAL	ID3	UTS2	CELA2B	SLC5A2	ATG10	MUC20	HIC1
47	0.01016	0.93995	MUC20	CAA	SLC5A2	THAL	BRAAK	ID3	APOE	HIC1	CELA2B	LYG2	SEVERITY
48	0.01016	0.95011	MUC20	CAA	SLC5A2	THAL	BRAAK	ID3	APOE	HIC1	CELA2B	LYG2	SEVERITY
49	0.01012	0.96023	MUC20	CAA	SLC5A2	THAL	BRAAK	ID3	APOE	HIC1	TUBA1C	CELA2B	DNAJC25
50	0.01011	0.97034	MUC20	CAA	SLC5A2	THAL	BRAAK	ID3	APOE	HIC1	TUBA1C	CELA2B	DNAJC25
51	0.00998	0.98032	MUC20	CAA	SLC5A2	THAL	BRAAK	ID3	APOE	HIC1	TUBA1C	CELA2B	DNAJC25
52	0.00994	0.99026	MUC20	CAA	SLC5A2	THAL	BRAAK	ID3	APOE	HIC1	TUBA1C	CELA2B	DNAJC25
53	0.00974	1.00000	MUC20	CAA	SLC5A2	THAL	BRAAK	ID3	HIC1	IL9	DNAJC25	LYG2	NRF1
Order rank	12	13	14	15	16	17	18	19	20	21	22	23	24
1	UTS2	MFSD2B	LYG2	ATG10	APOE4	ZNF135	U2AF1	STAB2	IL9	APOE	HIC1	SLC5A2	SEVERITY
2	HIC1	MFSD2B	LYG2	ATG10	APOE4	IL9	U2AF1	STAB2	ZNF135	APOE	UTS2	SLC5A2	SEVERITY
3	HIC1	MFSD2B	LYG2	ATG10	APOE4	IL9	U2AF1	STAB2	ZNF135	APOE	UTS2	SEVERITY	SLC5A2
4	HIC1	MFSD2B	LYG2	ATG10	APOE4	ZNF135	U2AF1	STAB2	IL9	APOE	UTS2	SLC5A2	SEVERITY
5	LYG2	TLR9	ID1	IL9	APOE	SLC5A2	ZNF135	MFSD2B	APOE4	SEVERITY	UTS2	HIC1	STAB2
6	LYG2	TLR9	ID1	IL9	APOE	SLC5A2	ZNF135	MFSD2B	APOE4	SEVERITY	UTS2	HIC1	STAB2
7	LYG2	ATG10	SLC5A2	IL9	MUC20	MFSD2B	ZNF135	ID1	APOE4	SEVERITY	UTS2	HIC1	STAB2
8	LYG2	ATG10	ID1	IL9	MUC20	MFSD2B	ZNF135	SLC5A2	APOE4	SEVERITY	UTS2	HIC1	STAB2
9	LYG2	ATG10	ID1	IL9	MUC20	SLC5A2	ZNF135	MFSD2B	APOE4	SEVERITY	UTS2	HIC1	STAB2
10	ATG10	UTS2	U2AF1	LYG2	ID1	NRF1	APOE	ZNF135	APOE4	STAB2	HIC1	TLR9	MUC20
11	LYG2	TLR9	ID1	IL9	MUC20	SLC5A2	ZNF135	MFSD2B	APOE4	SEVERITY	UTS2	HIC1	STAB2
12	NRF1	ZNF135	TLR9	ID1	IL9	APOE4	SLC5A2	LYG2	MFSD2B	APOE	SEVERITY	UTS2	HIC1
13	NRF1	LYG2	TLR9	ID1	IL9	APOE	SLC5A2	ZNF135	MFSD2B	APOE4	SEVERITY	UTS2	HIC1
14	NRF1	LYG2	TLR9	ID1	IL9	APOE4	SLC5A2	ZNF135	MFSD2B	APOE	SEVERITY	UTS2	HIC1
15	NRF1	LYG2	TLR9	ID1	IL9	APOE	SLC5A2	ZNF135	MFSD2B	APOE4	SEVERITY	UTS2	HIC1
16	UTS2	TLR9	MFSD2B	LYG2	ATG10	ATG10	ZNF135	U2AF1	STAB2	IL9	APOE	HIC1	SLC5A2
17	TLR9	UTS2	MFSD2B	LYG2	ATG10	APOE4	ZNF135	U2AF1	STAB2	IL9	APOE	HIC1	SLC5A2
18	NRF1	LYG2	TLR9	ID1	IL9	APOE	SLC5A2	ZNF135	MFSD2B	APOE4	SEVERITY	UTS2	HIC1

Table 13 (continued)

Order rank	12	13	14	15	16	17	18	19	20	21	22	23	24
19	ATG10	UTS2	ID1	IL9	U2AF1	NRF1	HIC1	ZNF135	APOE4	STAB2	APOE	TLR9	MUC20
20	NRF1	ZNF135	TLR9	ID1	IL9	APOE4	APOE	LYG2	MFSD2B	SLC5A2	SEVERITY	UTS2	HIC1
21	ATG10	UTS2	ID1	IL9	U2AF1	NRF1	HIC1	ZNF135	APOE4	STAB2	APOE	TLR9	MUC20
22	ATG10	UTS2	U2AF1	IL9	ID1	NRF1	HIC1	ZNF135	APOE4	STAB2	APOE	TLR9	MUC20
23	ATG10	UTS2	ID1	IL9	U2AF1	NRF1	HIC1	ZNF135	APOE4	SEVERITY	APOE	TLR9	MUC20
24	ZNF135	MFSD2B	STAB2	MUC20	APOE4	CELA2B	ID1	SEVERITY	SLC5A2	TUBA1C	LYG2	IL9	U2AF1
25	ZNF135	MFSD2B	STAB2	MUC20	SEVERITY	TLR9	ID1	APOE4	SLC5A2	TUBA1C	LYG2	IL9	U2AF1
26	ZNF135	MFSD2B	STAB2	MUC20	APOE4	TLR9	ID1	SEVERITY	SLC5A2	TUBA1C	LYG2	IL9	U2AF1
27	ATG10	UTS2	U2AF1	IL9	ID1	NRF1	APOE	ZNF135	APOE4	STAB2	HIC1	TLR9	MUC20
28	NRF1	ZNF135	TLR9	ID1	IL9	APOE4	APOE	LYG2	MFSD2B	U2AF1	SEVERITY	UTS2	HIC1
29	ATG10	UTS2	U2AF1	LYG2	ID1	NRF1	APOE	ZNF135	APOE4	STAB2	MFSD2B	TLR9	MUC20
30	NRF1	ZNF135	APOE	ID1	IL9	APOE4	TLR9	LYG2	MFSD2B	U2AF1	SEVERITY	UTS2	HIC1
31	UTS2	ID1	DNAJC25	TUBA1C	NRF1	ATG10	U2AF1	APOE4	ZNF135	STAB2	MFSD2B	TLR9	MUC20
32	UTS2	ATG10	DNAJC25	TUBA1C	NRF1	ID1	U2AF1	APOE4	ZNF135	STAB2	MFSD2B	TLR9	MUC20
33	UTS2	ATG10	IL9	TUBA1C	NRF1	ID1	U2AF1	APOE4	ZNF135	STAB2	MFSD2B	TLR9	MUC20
34	ATG10	UTS2	U2AF1	LYG2	ID1	NRF1	IL9	ZNF135	APOE4	STAB2	MFSD2B	TLR9	MUC20
35	NRF1	ZNF135	APOE	ID1	IL9	APOE4	TLR9	LYG2	MFSD2B	U2AF1	SEVERITY	UTS2	HIC1
36	ATG10	UTS2	U2AF1	IL9	ID1	NRF1	APOE	ZNF135	APOE4	STAB2	HIC1	TLR9	MUC20
37	APOE	ID1	ZNF135	TUBA1C	SEVERITY	U2AF1	ATG10	APOE4	CELA2B	STAB2	MFSD2B	TLR9	MUC20
38	APOE	ID1	CELA2B	TUBA1C	SEVERITY	U2AF1	ATG10	APOE4	ZNF135	STAB2	MFSD2B	TLR9	MUC20
39	APOE	ID1	CELA2B	TUBA1C	NRF1	U2AF1	ATG10	APOE4	ZNF135	STAB2	MFSD2B	TLR9	MUC20
40	LYG2	ATG10	SLC5A2	IL9	MUC20	MFSD2B	ZNF135	ID1	APOE4	UTS2	SEVERITY	HIC1	STAB2
41	LYG2	ATG10	SLC5A2	IL9	MUC20	MUC20	TLR9	LYG2	APOE4	SEVERITY	UTS2	HIC1	STAB2
42	NRF1	ZNF135	APOE	ID1	IL9	APOE4	TLR9	LYG2	MFSD2B	U2AF1	SEVERITY	MUC20	HIC1
43	NRF1	ZNF135	APOE	U2AF1	IL9	APOE4	TLR9	LYG2	MFSD2B	ID1	SEVERITY	MUC20	HIC1
44	LYG2	ATG10	ID1	IL9	MUC20	MFSD2B	ZNF135	SLC5A2	APOE4	SEVERITY	UTS2	HIC1	STAB2
45	MUC20	MFSD2B	DNAJC25	TUBA1C	NRF1	ID1	TLR9	ZNF135	APOE4	SEVERITY	IL9	STAB2	APOE4
46	LYG2	MFSD2B	DNAJC25	TUBA1C	NRF1	ID1	TLR9	ZNF135	U2AF1	SEVERITY	IL9	STAB2	APOE4
47	UTS2	IL9	ID1	DNAJC25	TUBA1C	NRF1	ATG10	U2AF1	APOE4	SEVERITY	IL9	STAB2	APOE4
48	IL9	UTS2	ID1	DNAJC25	TUBA1C	NRF1	ATG10	U2AF1	APOE4	ZNF135	STAB2	MFSD2B	TLR9
49	SEVERITY	ATG10	UTS2	U2AF1	LYG2	ID1	NRF1	IL9	APOE4	ZNF135	STAB2	MFSD2B	TLR9
50	SEVERITY	ATG10	UTS2	U2AF1	LYG2	ID1	NRF1	IL9	APOE4	ZNF135	STAB2	MFSD2B	TLR9
51	SEVERITY	ATG10	UTS2	IL9	LYG2	ID1	NRF1	U2AF1	APOE4	ZNF135	STAB2	MFSD2B	TLR9
52	SEVERITY	ATG10	UTS2	IL9	LYG2	NRF1	ID1	U2AF1	APOE4	ZNF135	STAB2	MFSD2B	TLR9
53	UTS2	APOE	ID1	ZNF135	TUBA1C	SEVERITY	MFSD2B	ATG10	APOE4	CELA2B	STAB2	U2AF1	TLR9

expression associated with APOE4 as well as lower NRF1 expression associated with APOE4. The association of the risk factor APOE4 with both ID3 and NRF1 expression in women suggest that these transcription regulators may exacerbate risk of CAA.

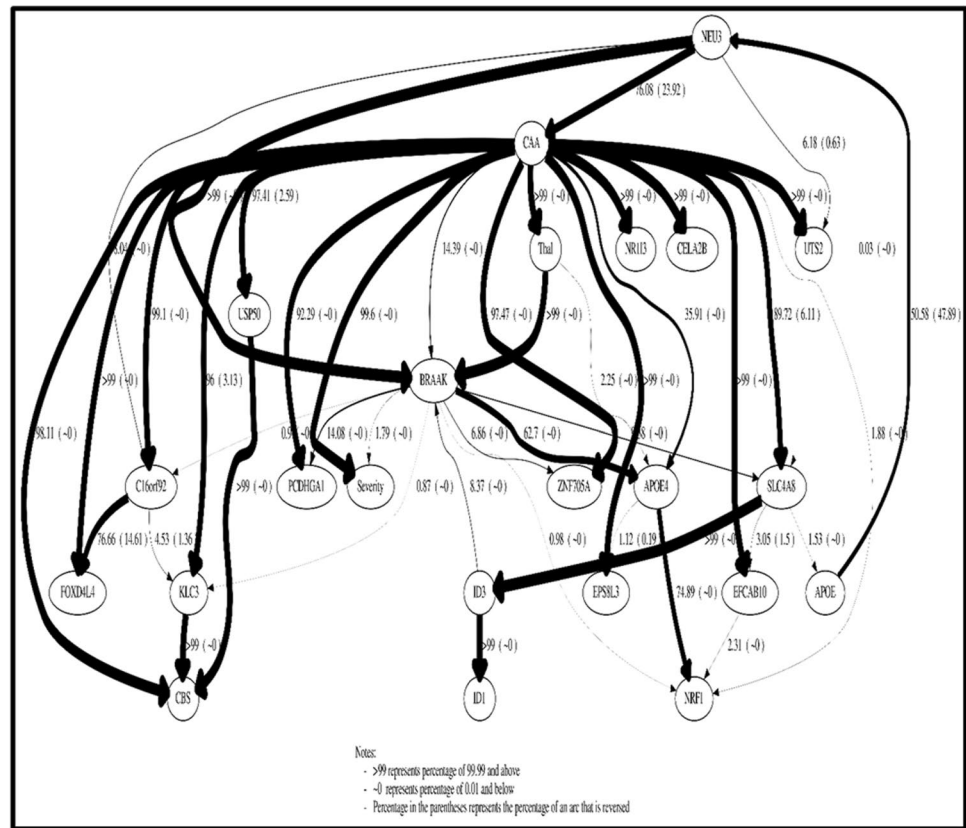
Genome wide transcriptional activity of both ID3 and NRF1 paralleled their mRNA expression levels. Total ID3 transcriptional activity was significantly increased in both CAA and AD patients as well as endothelial specific transcriptional activity. We also demonstrate ID3 activity is correlated with that of other transcription factors including SRF, STAT3, MZF1, CEBPA, E47, STAT5A, SP1, STAT1, SRF, IRF1, FOXJ2, IRF7, HSF1, SOX9, ATF6, FOXO3 (Fig. 9). We also discovered that NRF1 transcriptional activity was decreased in both CAA and AD patients. NRF1 activity correlated with other transcription factors including FOXO1, STAT3, MZF1, CEBPA, E47, STAT5A, SP1, STAT1, SRF, IRF1, FOXJ2, IRF7, HSF1, SOX9, ATF6, POU6F1, HSF2, PAX6, GFI1, MAX, FOXO3 and in AD including GFI1 and MAX transcription factors (Fig. 9). Gene pathways significantly downregulated in NRF1 total activity included endothelial cell signature, VEGF signaling, endothelial cell angiogenesis and vasculogenesis, and Wnt Signaling Pathway. Downregulation of angiogenesis and VEGF signaling could explain how NRF1 mediates injury of cerebral blood vessels in CAA. Therefore, we further examined sex specific differences in transcriptional gene signatures of both ID3 and NRF1 using Bayesian machine learning and the GeNIe simulation models.

Probabilistic graphical models were generated by machine learning based on 9,284 ID3 and 11,280 NRF1-bound ChIP-seq gene targets. ID3 and NRF1 targets were analyzed using a machine based causal Bayesian network search on RNA-Seq data from CAA patients. In ID3 causal BN analysis, we identified STAB2, MUC20, TUBA1C, HIC1, ZNF135, IL9, LYG2, MFSD2B, DNAJC25, CELA2B, TLR9, U2AF1, ATG10, UTS2, and SLC5A2 as Markov Blanket ID3 target genes associated with CAA (Fig. 14). In the overall analysis of the ID3 causal Bayesian network involving CAA, we discovered that ID3 was a direct child of CAA representing a significant association between the transcription factor and disease status in the cohort. Furthermore, modification of ID3 expression to a state of overexpression in GENIE analysis increased the marginal probability of CAA risk from 50 to 91% (Fig. 14a). Utilizing GENIE sensitivity analysis and combinatorial analysis resulted in the finding that increased ID3 combined with increased ATG10, SLC5A2, U2AF1, and UTS2, increased the risk of CAA in women patients by greater than 1.36×10^6 fold when compared to a gene expression pattern that showed normal or downregulated expression of these same genes. Seven different gene combination patterns of these genes consisting of ID3 overexpression versus downregulation resulted

in a significant increase in relative risk for CAA in females (Table 13). These results suggest that ID3 overexpression drives aberrant expression of ATG10, SLC5A2, U2AF1, and UTS2 in a manner that contributes to disease progression of CAA at the temporal cortex. ATG10 expression is necessary for the initiation of autophagy a process that is signified by the disassembling of damaged organelles from a cell during early stages of disease progression [38]. SLC5A2 is involved in hypoglycemic response, a clinical characteristic of increased risk in patients developing Alzheimer's Disease [39]. U2AF1 is a critical component of pre-mRNA splicing of stem/progenitor cells and results suggest that it may have been overexpressed in our cohort in an effort to adapt to a loss of cells as CAA progressed in patients [40]. Human population studies have identified UTS2 polymorphisms as a significant factor contributing to migraines, and individuals with migraines are three times more likely to develop dementia [41, 42]. In terms of genes that were identified to be contributing to the ID3 representative network, we found gene candidates with prior knowledge of involvement in CAA pathology. SLC5A2, UTS2, DNAJC25, and U2AF1 are all involved in processes contributing to CAA pathogenesis including vascular hypertension, angiogenesis, maintenance of blood pressure and response to stress factors [41–44].

In NRF1 causal BN analysis, we identified ID3, PXMP4, SERHL, EXOSC6, MCM3AP-AS1, ZNF570, RPL39, SNORA31, TRIM44, SNORA64, SATL1, ZNF135, TOMM6, SLC9B1, SNORA33, GATC, and GDF9 as Markov Blanket NRF1 motif enriched genes associated with CAA (Fig. 15). In the overall analysis of the NRF1 causal Bayesian network involving CAA, we discovered a direct connection between CAA $\rightarrow$ Thal $\rightarrow$ NRF1 resulting in a significant association between NRF1 expression and Thal phase a marker of amyloid beta deposition in cerebral microvessels. When examining how the modification of NRF1 expression impacted the probability of CAA in our cohort, we found that the downregulation of NRF1 expression resulted in probability of CAA increasing to 85% from 49% probability of CAA (Fig. 15A). When inputted into GENIE sensitivity analysis and combinatorial analysis resulted in the finding that decreased NRF1 combined with normal or decreased expression of GATC/SLC9B1 and increased expression of GDF9, TOMM6, ZNF135 resulted in increased risk of CAA in male patients by greater than 8.20×10^7 fold when compared to a gene expression pattern that exhibited opposing patterns of expression for all the aforementioned genes. Furthermore, seven different gene combination patterns of these genes consisting of NRF1 downregulation versus NRF1 overexpression resulted in a significant increase in relative risk for CAA in males (Table 12). These results suggest that NRF1 downregulation potentially serves as a driver of modified expression of

Fig. 17 MCMC ordered ID3 network for men with CAA



GATC, SLC9B1, GDF9, TOMM6, and ZNF135 that potentially drives CAA progression in men. GATC is a target of NRF1 that is involved in glutamyl-tRNAIn biosynthesis transamidation and mitochondrial translation, being implicated in mitochondrial disorders significantly associated with increased risk for Alzheimer's Disease [45, 46]. GDF9 is a key component of the TGF β signaling pathway involved in the inhibition of vascular endothelial growth factor (VEGF) key to pro-angiogenic activity [47]. SNORA33 directly interacts with EZH2 to promote epigenetic reprogramming in cardiac hypertrophy and may be involved in epigenetic interactions that lead to increased aging of brain tissue associated with Alzheimer disease [48, 49]. TOMM6 regulates protein translocase involved in mitochondrial survival and wellbeing suggesting that its increased expression was a compensatory mechanism to maintain mitochondrial levels of patients with CAA [50]. Homozygous transition mutation of ZNF135 is associated with neurological and developmental symptoms in a transgenic mouse line and are results suggest it may play a role in the pathogenesis of CAA [51]. SLC9B1 gene expression is regulated by DNA methylation-dependent and independent mechanisms mediated by several DNA regulatory elements and is linked with Christianson syndrome a disease that affects intellectual disability leading to symptoms of dementia [52, 53]. Based on these findings, this NRF1 network may contribute to the

progression of CAA progression. Our study used publicly available RNA-seq data from CAA patients; however, it is important to validate our findings in future studies by testing new CAA patient samples. This will give us the opportunity to determine if our transcriptional signatures can effectively distinguish a CAA patient from a healthy control. By employing the models showcased in Figs. 14 and 15, we can generate predictions for each new case and assess the predictive capabilities of each model. For instance, we can report the Receiver Operating Curve (ROC) to quantify the performance of each model. A higher area under the ROC (> 0.70) indicates a stronger predictive power for the model. Furthermore, it is worth mentioning that the new CAA samples can also contribute to enhancing the models for more comprehensive causal analysis through the utilization of MCMC order search.

Monte Carlo Markov Chain (MCMC) gene ordering approach showed sex-specific causal transcriptional gene signatures driven by ID3 and NRF1 in CAA. The majority of MCMC gene orders found in Table 13 show that ID3 drives severity of CAA (ID3 $\rightarrow$ APOE4 $\rightarrow$ severity) in women. In men, NRF1 was the causal parent of CAA and severity in 100% of all gene orders. These results in Table 16 show that NRF1 drives CAA (NRF1 $\rightarrow$ CAA $\rightarrow$ severity). Additionally, we discovered that DNAJC25, APOE, STAB2, and MUC20 targets of

Table 14 Markov blanket MCMC causal gene orders of CAA male cases/controls and ID3 targets

Order rank	Order probability	Cumulative order probability	1	2	3	4	5	6	7	8	9	10	11
1	0.04480	0.04480	APOE	NEU3	CAA	C16orf92	FOXD4L4	Thal	NR1I3	KLC3	USP50	BRAAK	SLC4A8
2	0.04181	0.08662	APOE	NEU3	CAA	C16orf92	FOXD4L4	KLC3	NR1I3	Thal	USP50	BRAAK	SLC4A8
3	0.03979	0.12640	APOE	NEU3	CAA	C16orf92	FOXD4L4	KLC3	NR1I3	USP50	Thal	BRAAK	SLC4A8
4	0.03768	0.16409	CAA	UTS2	Thal	NEU3	BRAAK	KLC3	CELA2B	EFCAB10	APOE4	SLC4A8	NR1I3
5	0.03565	0.19974	APOE	NEU3	CAA	C16orf92	FOXD4L4	Thal	Severity	KLC3	USP50	BRAAK	SLC4A8
6	0.03510	0.23484	APOE	NEU3	CAA	C16orf92	FOXD4L4	Thal	Severity	EFCAB10	USP50	BRAAK	SLC4A8
7	0.03495	0.26979	APOE	NEU3	CAA	C16orf92	FOXD4L4	Thal	Severity	EFCAB10	USP50	BRAAK	SLC4A8
8	0.03150	0.30128	CAA	SLC4A8	Thal	NEU3	BRAAK	KLC3	CELA2B	EFCAB10	APOE4	UTS2	C16orf92
9	0.03073	0.33202	CAA	SLC4A8	Thal	NEU3	BRAAK	KLC3	CELA2B	EFCAB10	APOE4	UTS2	C16orf92
10	0.03055	0.36256	SLC4A8	ID3	CAA	UTS2	NEU3	FOXD4L4	Thal	EPS8L3	USP50	BRAAK	APOE
11	0.03055	0.39311	SLC4A8	ID3	CAA	UTS2	NEU3	FOXD4L4	Thal	EPS8L3	USP50	BRAAK	APOE
12	0.02524	0.41835	NEU3	CAA	EFCAB10	C16orf92	Thal	FOXD4L4	USP50	NR1I3	BRAAK	NR1I	EPS8L3
13	0.02418	0.44253	NEU3	CAA	C16orf92	FOXD4L4	KLC3	NR1I3	USP50	Thal	BRAAK	SLC4A8	UTS2
14	0.02410	0.46662	NEU3	CAA	C16orf92	FOXD4L4	KLC3	NR1I3	USP50	Thal	BRAAK	SLC4A8	UTS2
15	0.02387	0.49049	APOE	NEU3	CAA	C16orf92	FOXD4L4	Thal	ZNF705A	EFCAB10	USP50	BRAAK	SLC4A8
16	0.02187	0.51236	NEU3	CAA	EFCAB10	C16orf92	Thal	NR1I	USP50	NR1I3	BRAAK	FOXD4L4	EPS8L3
17	0.02187	0.53423	NEU3	CAA	EFCAB10	C16orf92	Thal	NR1I	USP50	NR1I3	BRAAK	EPS8L3	FOXD4L4
18	0.02100	0.55523	CAA	SLC4A8	Thal	NEU3	BRAAK	KLC3	CELA2B	EFCAB10	APOE4	UTS2	NR1I3
19	0.02088	0.57611	NEU3	CAA	C16orf92	FOXD4L4	CELA2B	APOE4	NR1I	Thal	BRAAK	Severity	SLC4A8
20	0.01722	0.59333	APOE	NEU3	CAA	C16orf92	FOXD4L4	KLC3	NR1I3	USP50	Thal	BRAAK	SLC4A8
21	0.01721	0.61054	APOE	NEU3	CAA	C16orf92	FOXD4L4	KLC3	NR1I3	USP50	Thal	BRAAK	SLC4A8
22	0.01615	0.62670	NEU3	CAA	C16orf92	FOXD4L4	KLC3	NR1I3	USP50	Thal	BRAAK	SLC4A8	UTS2
23	0.01606	0.64275	NEU3	CAA	C16orf92	FOXD4L4	KLC3	NR1I3	USP50	Thal	BRAAK	SLC4A8	UTS2
24	0.01572	0.65847	APOE	NEU3	CAA	C16orf92	FOXD4L4	EFCAB10	APOE4	NR1I	Thal	BRAAK	USP50
25	0.01559	0.67407	APOE	NEU3	CAA	C16orf92	FOXD4L4	NR1I3	APOE4	NR1I	Thal	BRAAK	USP50
26	0.01557	0.68963	APOE	NEU3	CAA	C16orf92	FOXD4L4	NR1I3	APOE4	NR1I	Thal	BRAAK	Severity
27	0.01556	0.70519	APOE	NEU3	CAA	C16orf92	FOXD4L4	NR1I3	APOE4	NR1I	Thal	BRAAK	USP50
28	0.01552	0.72071	APOE	NEU3	CAA	C16orf92	FOXD4L4	NR1I3	APOE4	NR1I	Thal	BRAAK	Severity
29	0.01550	0.73621	NEU3	CAA	APOE4	FOXD4L4	CELA2B	C16orf92	NR1I	Thal	BRAAK	KLC3	SLC4A8
30	0.01549	0.75170	NEU3	CAA	APOE4	FOXD4L4	CELA2B	C16orf92	NR1I	Thal	BRAAK	KLC3	SLC4A8
31	0.01542	0.76712	NEU3	CAA	APOE4	FOXD4L4	CELA2B	C16orf92	NR1I	Thal	BRAAK	Severity	SLC4A8
32	0.01530	0.78242	APOE	NEU3	CAA	C16orf92	FOXD4L4	EFCAB10	APOE4	NR1I	Thal	BRAAK	USP50
33	0.01528	0.79770	APOE	NEU3	CAA	C16orf92	FOXD4L4	EFCAB10	APOE4	NR1I	Thal	BRAAK	USP50
34	0.01515	0.81285	APOE	NEU3	CAA	C16orf92	FOXD4L4	EFCAB10	APOE4	EPS8L3	Thal	BRAAK	USP50
35	0.01487	0.82772	APOE	NEU3	CAA	C16orf92	FOXD4L4	CELA2B	APOE4	NR1I	Thal	BRAAK	Severity
36	0.01486	0.84258	APOE	NEU3	CAA	C16orf92	FOXD4L4	CELA2B	APOE4	NR1I	Thal	BRAAK	Severity

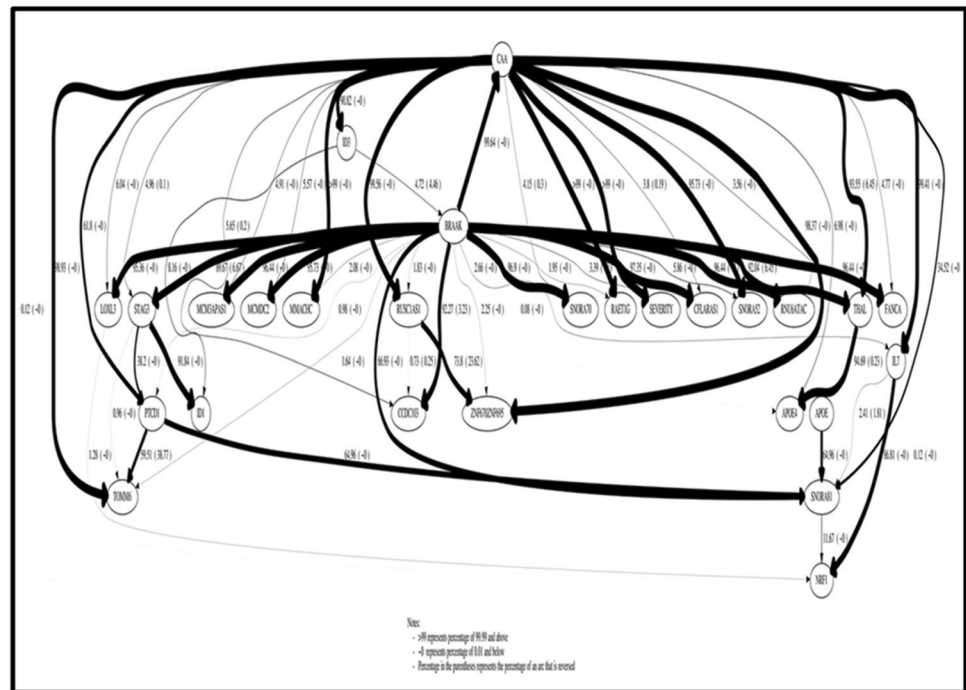
Table 14 (continued)

Order rank	Order probability	Cumulative order probability	1	2	3	4	5	6	7	8	9	10	11
37	0.01404	0.85662	APOE	NEU3	CAA	C16orf92	FOXDL4	FOXDL3	EFCAB10	UTS2	EPS8L3	Thal	USP50
38	0.01382	0.87043	APOE	NEU3	CAA	C16orf92	FOXDL3	EPS8L3	NR1I3	APOE4	NRF1	Thal	Severity
39	0.01306	0.88350	NEU3	CAA	C16orf92	FOXDL4	KLC3	EPS8L3	NR1I3	Severity	Thal	BRAAK	UTS2
40	0.01296	0.89646	USP50	CAA	NEU3	C16orf92	FOXDL3	EPS8L3	Thal	ZNF705A	EFCAB10	SLC4A8	NR1I3
41	0.01289	0.90935	USP50	CAA	NEU3	C16orf92	ZNF705A	Thal	Thal	EPS8L3	EFCAB10	SLC4A8	NR1I3
42	0.01262	0.92197	NEU3	CAA	APOE4	NRF1	Thal	Thal	ZNF705A	NR1I3	USP50	BRAAK	KLC3
43	0.01260	0.93457	NEU3	CAA	APOE4	NRF1	Thal	Thal	ZNF705A	NR1I3	USP50	BRAAK	KLC3
44	0.01143	0.94600	APOE	NEU3	CAA	C16orf92	FOXDL4	FOXDL3	Thal	ZNF705A	PCDHGA1	USP50	SLC4A8
45	0.01142	0.95742	APOE	NEU3	CAA	C16orf92	FOXDL4	FOXDL3	Thal	ZNF705A	PCDHGA1	USP50	SLC4A8
46	0.01124	0.96866	APOE	NEU3	CAA	CELA2B	FOXDL4	FOXDL3	C16orf92	APOE4	NRF1	Thal	Severity
47	0.01071	0.97937	KLC3	CAA	NEU3	C16orf92	EPS8L3	EPS8L3	Thal	ZNF705A	EFCAB10	SLC4A8	NR1I3
48	0.01071	0.99008	KLC3	CAA	NEU3	C16orf92	EPS8L3	EPS8L3	Thal	ZNF705A	EFCAB10	SLC4A8	NR1I3
49	0.00992	1.00000	KLC3	CAA	NEU3	C16orf92	EPS8L3	EPS8L3	Thal	ZNF705A	EFCAB10	SLC4A8	NR1I3
Order rank	12	13	14	15	16	17	18	19	20	21	22	23	
1	UTS2	EFCAB10	PCDHGA1	APOE4	ID3	NRF1	Severity	ID1	EPS8L3	ZNF705A	CBS	CELA2B	
2	UTS2	EFCAB10	PCDHGA1	APOE4	ID3	NRF1	Severity	ID1	EPS8L3	ZNF705A	CBS	CELA2B	
3	UTS2	EFCAB10	PCDHGA1	APOE4	ID3	NRF1	Severity	ID1	EPS8L3	ZNF705A	CBS	CELA2B	
4	FOXDL4	ID3	APOE	USP50	Severity	NRF1	C16orf92	PCDHGA1	CBS	ZNF705A	EPS8L3	ID1	
5	UTS2	EFCAB10	PCDHGA1	APOE4	ID3	NRF1	NR1I3	ID1	EPS8L3	ZNF705A	CBS	CELA2B	
6	ZNF705A	KLC3	PCDHGA1	APOE4	ID3	NRF1	NR1I3	ID1	EPS8L3	UTS2	CBS	CELA2B	
7	UTS2	KLC3	PCDHGA1	APOE4	ID3	NRF1	NR1I3	ID1	EPS8L3	ZNF705A	CBS	CELA2B	
8	FOXDL4	ID3	APOE	USP50	Severity	NRF1	NR1I3	ID1	EPS8L3	ZNF705A	EPS8L3	ID1	
9	FOXDL4	ZNF705A	APOE	USP50	Severity	NRF1	NR1I3	ID1	EPS8L3	ZNF705A	EPS8L3	ID1	
10	ID1	KLC3	CBS	PCDHGA1	CELA2B	NR1I3	Severity	C16orf92	EFCAB10	ZNF705A	APOE4	NRF1	
11	ID1	KLC3	CBS	PCDHGA1	NR1I3	CELA2B	Severity	C16orf92	EFCAB10	ZNF705A	APOE4	NRF1	
12	KLC3	SLC4A8	ZNF705A	ID3	PCDHGA1	ID1	Severity	CELA2B	APOE	CBS	APOE4	UTS2	
13	EFCAB10	NRF1	APOE4	PCDHGA1	ID3	Severity	ID1	EPS8L3	ZNF705A	CBS	CELA2B	APOE	
14	EFCAB10	NRF1	APOE4	PCDHGA1	ID3	Severity	CELA2B	EPS8L3	ZNF705A	CBS	ID1	APOE	
15	Severity	KLC3	PCDHGA1	APOE4	ID3	NRF1	NR1I3	ID1	EPS8L3	UTS2	CBS	CELA2B	
16	KLC3	SLC4A8	ZNF705A	ID3	PCDHGA1	ID1	Severity	CELA2B	APOE	CBS	APOE4	UTS2	
17	KLC3	SLC4A8	ZNF705A	ID3	PCDHGA1	ID1	Severity	CELA2B	APOE	CBS	APOE4	UTS2	
18	FOXDL4	ID3	APOE	USP50	Severity	NRF1	C16orf92	PCDHGA1	CBS	ZNF705A	EPS8L3	ID1	
19	USP50	EPS8L3	KLC3	PCDHGA1	ID3	ZNF705A	NR1I3	UTS2	EFCAB10	CBS	ID1	APOE	
20	UTS2	EFCAB10	NRF1	APOE4	ID3	PCDHGA1	Severity	ID1	EPS8L3	ZNF705A	CBS	CELA2B	

Table 14 (continued)

Order rank	12	13	14	15	16	17	18	19	20	21	22	23
21	UTS2	EFCAB10	NRF1	APOE4	PCDHGA1	ID3	Severity	ID1	EPS8L3	ZNF705A	CBS	CELA2B
22	EPS8L3	NRF1	APOE4	PCDHGA1	ID3	EFCAB10	CELA2B	Severity	ZNF705A	CBS	ID1	APOE
23	EPS8L3	NRF1	APOE4	PCDHGA1	ID3	Severity	CELA2B	EFCAB10	ZNF705A	CBS	ID1	APOE
24	SLC4A8	KLC3	EPS8L3	UTS2	PCDHGA1	ID3	NR1I3	CELA2B	Severity	ZNF705A	CBS	ID1
25	SLC4A8	Severity	EFCAB10	KLC3	PCDHGA1	ID3	EPS8L3	CELA2B	UTS2	ZNF705A	CBS	ID1
26	SLC4A8	USP50	EFCAB10	KLC3	PCDHGA1	ID3	EPS8L3	CELA2B	UTS2	ZNF705A	CBS	ID1
27	SLC4A8	Severity	EPS8L3	KLC3	PCDHGA1	ID3	EFCAB10	CELA2B	UTS2	ZNF705A	CBS	ID1
28	SLC4A8	USP50	ZNF705A	KLC3	PCDHGA1	ID3	EPS8L3	CELA2B	UTS2	EFCAB10	CBS	ID1
29	USP50	EPS8L3	Severity	PCDHGA1	ID3	ZNF705A	NR1I3	UTS2	CBS	EFCAB10	ID1	APOE
30	USP50	EPS8L3	Severity	PCDHGA1	ID3	ZNF705A	NR1I3	UTS2	EFCAB10	CBS	ID1	APOE
31	USP50	EPS8L3	KLC3	PCDHGA1	ID3	ZNF705A	NR1I3	UTS2	EFCAB10	CBS	ID1	APOE
32	SLC4A8	Severity	EPS8L3	UTS2	PCDHGA1	ID3	NR1I3	CELA2B	KLC3	ZNF705A	CBS	ID1
33	SLC4A8	Severity	EPS8L3	KLC3	PCDHGA1	ID3	NR1I3	CELA2B	UTS2	ZNF705A	CBS	ID1
34	SLC4A8	KLC3	NRF1	UTS2	PCDHGA1	ID3	NR1I3	CELA2B	Severity	ZNF705A	CBS	ID1
35	SLC4A8	USP50	ZNF705A	KLC3	PCDHGA1	ID3	EPS8L3	NR1I3	UTS2	EFCAB10	CBS	ID1
36	SLC4A8	USP50	EPS8L3	KLC3	PCDHGA1	ID3	ZNF705A	NR1I3	UTS2	EFCAB10	CBS	ID1
37	SLC4A8	KLC3	NRF1	APOE4	PCDHGA1	ID3	NR1I3	CELA2B	Severity	ZNF705A	CBS	ID1
38	SLC4A8	USP50	EFCAB10	KLC3	PCDHGA1	ID3	FOXDL4	CELA2B	UTS2	ZNF705A	CBS	ID1
39	EPS8L3	NRF1	APOE4	PCDHGA1	ID3	EFCAB10	CELA2B	USP50	ZNF705A	CBS	ID1	APOE
40	APOE4	ID3	APOE	ID1	BRAAK	Severity	PCDHGA1	FOXDL4	NRF1	KLC3	CBS	CELA2B
41	APOE4	ID3	APOE	ID1	BRAAK	Severity	PCDHGA1	FOXDL4	NRF1	KLC3	CBS	CELA2B
42	SLC4A8	EFCAB10	CBS	ID3	PCDHGA1	C16orf92	EPS8L3	FOXDL4	UTS2	ID1	CELA2B	APOE
43	SLC4A8	EFCAB10	PCDHGA1	ID3	CBS	C16orf92	EPS8L3	FOXDL4	UTS2	ID1	CELA2B	APOE
44	Severity	KLC3	EFCAB10	APOE4	ID3	NRF1	NR1I3	ID1	CBS	UTS2	EPS8L3	CELA2B
45	Severity	KLC3	EFCAB10	APOE4	ID3	NRF1	NR1I3	ID1	EPS8L3	UTS2	CBS	CELA2B
46	SLC4A8	USP50	ZNF705A	KLC3	PCDHGA1	ID3	EPS8L3	NR1I3	UTS2	EFCAB10	CBS	ID1
47	APOE4	ID3	APOE	ID1	UTS2	Severity	FOXDL4	PCDHGA1	NRF1	USP50	CBS	CELA2B
48	APOE4	ID3	APOE	ID1	UTS2	Severity	PCDHGA1	FOXDL4	NRF1	USP50	CBS	CELA2B
49	APOE4	ID3	APOE	ID1	BRAAK	Severity	PCDHGA1	FOXDL4	NRF1	USP50	CBS	CELA2B

Fig. 18 MCMC ordered NRF1 network for women with CAA



ID3 played a significant role in causality for women with CAA. DNAJC25 is a member of the HSP40/JDP subfamily C found to be associated with p53 as a heat shock protein response that when overexpressed is shown to significantly increase cell apoptosis a hallmark of CAA [54]. The discovery of APOE as a parent of CAA and ID3 target is significant given the established role of the APOE gene in the accumulation of amyloid beta protein in the brain [55]. STAB2 is known to encode a protein that is associated with inflammation, overexpression of this gene as reported by our analysis as ID3 targeting may be contributing to cerebral amyloid inflammation a hallmark of CAA progression [55, 56]. MUC20 is a known inducer of angiogenesis through the activation of vascular endothelial growth factors suggesting that it is involved in pathological angiogenesis given our data [57]. In men with CAA, NRF1 regulated transcriptional gene signatures that were causal for CAA included MCM3AP1 and ID1. MCM3AP-AS1 is involved in RNA processing and implicated in several hallmarks of CAA such as cell proliferation, apoptosis, cell cycle, and cell migration [58]. ID1 promotes endothelial cell survival and regeneration, in addition the ID1/HIF-1 pathway is evidenced to contribute to cell cycle reentry in the in vitro AD model of primary cortical neurons [59]. Therefore, in both ID3 and NRF1 representative networks we report causal gene networks and gene expression patterns that contributed to cell cycle arrest, apoptosis, and senescence hallmarks that substantiate the involvement of these two transcription regulators to CAA disease progression.

In conclusion, dynamic machine learning of RNA-seq of CAA patients show that both ID3 and NRF1 were associated with CAA in a sex specific manner. Aberrant ID3 and NRF1 activity were associated with the pathogenesis and severity of CAA. NRF1 mRNA expression was significantly downregulated in both CAA and AD cases across two separate brain regions (cerebellum and temporal cortex) and across sex. Genome wide ID3 and NRF1 activity was significantly correlated to pathways that contribute to hallmarks of CAA such as: VEGF Signaling, cell apoptosis, and endothelial cell migration amongst others. Gene networks of ID3 targets and NRF1 motif enriched causal genes from DEG analysis were discovered through Bayesian network structure learning algorithm. The novel findings that have emerged from this study include ID3/ NRF1 causal transcriptional gene signatures associated with CAA and the significant influence of these signatures on CAA risk by modified gene expression of ID3/ NRF1 as well as gene combinations. MCMC gene ordering showed that NRF1 is a sex-specific driver of CAA in men while ID3 contributes to CAA severity in women. Taken together, our findings provide the first evidence that both the ID3 and NRF1 transcription gene signatures influence CAA risk. These results may provide clinical prognosis and therapeutic targeting of ID3 and NRF1 driven molecular signature differences across sex-specific CAA progression moving forward. Further analyses of ID3- and NRF1-regulated molecular drivers of CAA may provide new targets for personalized medicine and/or prevention strategies against CAA.

Table 15 Markov blanket MCMC causal gene orders of CAA female cases/controls and NRFI targets

Order ranking	Order probability	Cumulative order probability	1	2	3	4	5	6	7	8
1	0.064475178	0.064475178	THAL	BRAAK	CFLARAS1	CAA	SNORA52	IL7	MCMDC2	RNU6ATAC
2	0.034094896	0.098570074	BRAAK	CAA	THAL	SEVERITY	MCMDC2	RNU6ATAC	SNORA70	LOXL3
3	0.033056854	0.131626928	BRAAK	CAA	THAL	SEVERITY	RAETIG	RNU6ATAC	SNORA70	LOXL3
4	0.033052434	0.164679362	BRAAK	CAA	THAL	SEVERITY	RAETIG	RNU6ATAC	SNORA70	LOXL3
5	0.033047411	0.197726773	BRAAK	CAA	THAL	SEVERITY	MCMDC2	RNU6ATAC	SNORA70	LOXL3
6	0.032948485	0.230675259	BRAAK	CAA	THAL	SEVERITY	RAETIG	RNU6ATAC	SNORA70	LOXL3
7	0.031304803	0.261980062	BRAAK	CAA	ID3	THAL	MCM3APAS1	RNU6ATAC	SNORA70	APOE4
8	0.031285708	0.29326577	BRAAK	CAA	ID3	THAL	MCM3APAS1	MMACHC	SNORA70	APOE4
9	0.025328196	0.318593967	BRAAK	SNORA70	CAA	IL7	RAETIG	LOXL3	MMACHC	RNU6ATAC
10	0.023502955	0.342096922	BRAAK	CAA	ID3	THAL	CFLARAS1	MMACHC	SNORA81	FANCA
11	0.02346557	0.365562491	BRAAK	CAA	ID3	THAL	CFLARAS1	MMACHC	SNORA81	SNORA52
12	0.023424755	0.388987246	BRAAK	CAA	ID3	THAL	CFLARAS1	MMACHC	SNORA70	SNORA52
13	0.02099601	0.409983256	BRAAK	SNORA70	CAA	IL7	RAETIG	CFLARAS1	MMACHC	RNU6ATAC
14	0.01923161	0.429214866	BRAAK	CAA	ID3	SNORA52	MCM3APAS1	RNU6ATAC	SNORA70	LOXL3
15	0.019017037	0.448231903	BRAAK	CAA	ID3	SEVERITY	MCM3APAS1	RNU6ATAC	SNORA70	LOXL3
16	0.017650126	0.465882029	BRAAK	SNORA70	CAA	FANCA	MCM3APAS1	LOXL3	RNU6ATAC	SEVERITY
17	0.017552977	0.483435006	BRAAK	SNORA70	CAA	FANCA	MCM3APAS1	LOXL3	RNU6ATAC	SEVERITY
18	0.017178044	0.50061305	BRAAK	CAA	LOXL3	SEVERITY	MCMDC2	FANCA	MMACHC	RUSCIAS1
19	0.016880341	0.517493391	BRAAK	CAA	FANCA	SEVERITY	MCMDC2	LOXL3	MMACHC	RUSCIAS1
20	0.016878474	0.534371865	BRAAK	CAA	LOXL3	SEVERITY	MCMDC2	FANCA	MMACHC	RUSCIAS1
21	0.016724774	0.551096638	STAG3	BRAAK	CFLARAS1	CAA	SNORA52	IL7	MCMDC2	RNU6ATAC
22	0.016722686	0.567819325	BRAAK	SNORA70	CAA	FANCA	CFLARAS1	LOXL3	RNU6ATAC	SEVERITY
23	0.016715467	0.584534792	BRAAK	SNORA70	CAA	FANCA	CFLARAS1	LOXL3	RNU6ATAC	SEVERITY
24	0.016694656	0.601229448	STAG3	BRAAK	CFLARAS1	CAA	SNORA52	IL7	MCMDC2	RNU6ATAC
25	0.016678012	0.617907461	STAG3	BRAAK	CFLARAS1	CAA	SNORA52	IL7	MCMDC2	RNU6ATAC
26	0.016607624	0.634515085	STAG3	BRAAK	CFLARAS1	CAA	SNORA52	IL7	MCMDC2	RNU6ATAC
27	0.014790014	0.649305098	BRAAK	RNU6ATAC	CAA	APOE	RUSCIAS1	MMACHC	ID3	CCDC103
28	0.014208471	0.66351357	BRAAK	SNORA70	CAA	APOE	RNU6ATAC	CFLARAS1	RAETIG	SNORA52
29	0.014049158	0.677562728	ID3	BRAAK	CFLARAS1	CAA	SNORA52	IL7	MCMDC2	RNU6ATAC
30	0.013978444	0.691541172	BRAAK	CCDC103	CAA	PTCD1	SEVERITY	MCMDC2	FANCA	ZNF670ZNF695
31	0.013947643	0.705488815	BRAAK	SNORA70	CAA	FANCA	RNU6ATAC	CFLARAS1	RAETIG	SNORA52
32	0.013934444	0.719423259	BRAAK	CCDC103	CAA	PTCD1	SNORA70	MCMDC2	MCM3APAS1	ZNF670ZNF695
33	0.013782429	0.733205687	BRAAK	RNU6ATAC	CAA	APOE	RUSCIAS1	MMACHC	ID3	CCDC103
34	0.013725561	0.746931248	BRAAK	CCDC103	CAA	PTCD1	SEVERITY	MCMDC2	FANCA	ZNF670ZNF695
35	0.012852859	0.759784108	BRAAK	SNORA70	CAA	FANCA	RNU6ATAC	CFLARAS1	RAETIG	SNORA52
36	0.012842281	0.772626389	BRAAK	SNORA70	CAA	FANCA	RNU6ATAC	CFLARAS1	RAETIG	IL7
37	0.012692385	0.785318774	BRAAK	CCDC103	CAA	LOXL3	SEVERITY	MCMDC2	FANCA	ZNF670ZNF695
38	0.012639915	0.797958689	BRAAK	CCDC103	CAA	PTCD1	SEVERITY	MCMDC2	FANCA	ZNF670ZNF695

Table 15 (continued)

Order ranking	Order probability	Cumulative order prob-ability	1	2	3	4	5	6	7	8	
39	0.011952333	0.809911022	BRAAK	BRAAK	RNU6ATAC	CAA	APOE	MCM3APAS1	MMACHC	ID3	CCDC103
40	0.011589055	0.821500078	ID3	BRAAK	SNORA70	CAA	CAA	FANCA	MCM3APAS1	LOXL3	RNU6ATAC
41	0.011335592	0.83283567	BRAAK	BRAAK	CCDC103	CAA	LOXL3	SEVERITY	MCMDC2	FANCA	IL7
42	0.011195884	0.844031554	BRAAK	BRAAK	CCDC103	CAA	LOXL3	SEVERITY	MCMDC2	FANCA	IL7
43	0.011180372	0.855211925	BRAAK	BRAAK	CCDC103	CAA	LOXL3	SEVERITY	MCMDC2	FANCA	IL7
44	0.011166375	0.866378301	CCDC103	BRAAK	BRAAK	CAA	ID3	SEVERITY	MCM3APAS1	RNU6ATAC	SNORA70
45	0.011164583	0.877542883	CCDC103	BRAAK	BRAAK	CAA	ID3	SEVERITY	MCM3APAS1	RNU6ATAC	SNORA70
46	0.010762613	0.888305496	ID3	BRAAK	BRAAK	SNORA70	CAA	FANCA	MCM3APAS1	LOXL3	RNU6ATAC
47	0.010761419	0.899066915	ID3	BRAAK	BRAAK	SNORA70	CAA	FANCA	MCM3APAS1	LOXL3	RNU6ATAC
48	0.010445072	0.909511987	BRAAK	BRAAK	CAA	THAL	THAL	CFLARAS1	MMACHC	SNORA81	FANCA
49	0.010341076	0.919853064	BRAAK	BRAAK	CAA	MMACHC	THAL	CFLARAS1	ID3	SNORA81	ID1
50	0.010341058	0.930194122	BRAAK	BRAAK	CAA	MMACHC	THAL	CFLARAS1	ID3	SNORA81	FANCA
51	0.010226332	0.940420454	BRAAK	BRAAK	SNORA70	CAA	IL7	RAETIG	CFLARAS1	MMACHC	RNU6ATAC
52	0.010054992	0.950475446	BRAAK	BRAAK	CAA	ID3	THAL	CFLARAS1	MMACHC	SNORA81	FANCA
53	0.010035662	0.960511108	BRAAK	BRAAK	CAA	ID3	THAL	CFLARAS1	MMACHC	SNORA81	FANCA
54	0.01001358	0.970524688	CCDC103	BRAAK	BRAAK	CAA	LOXL3	SEVERITY	MCMDC2	FANCA	MMACHC
55	0.01001358	0.980538268	BRAAK	BRAAK	CCDC103	CAA	LOXL3	SEVERITY	MCMDC2	FANCA	MMACHC
56	0.009974126	0.990512393	BRAAK	BRAAK	CAA	ID3	THAL	CFLARAS1	ID1	SNORA81	FANCA
57	0.009487607	1	BRAAK	BRAAK	CCDC103	CAA	LOXL3	SEVERITY	MCMDC2	FANCA	MMACHC
Order ranking	9	10	11	12	13	14	15	16	17	18	
1	SEVERITY	TOMM6	FANCA	STAG3	ID3	PTCD1	SNORA70	APOE	RUSCIAS1	SNORA81	
2	PTCD1	ID3	MMACHC	FANCA	RUSCIAS1	APOE4	ZNF670ZNF695	MCM3APAS1	STAG3	IL7	
3	PTCD1	ID3	MMACHC	STAG3	RUSCIAS1	APOE	ZNF670ZNF695	MCM3APAS1	FANCA	IL7	
4	MMACHC	ID3	PTCD1	STAG3	RUSCIAS1	APOE	ZNF670ZNF695	MCM3APAS1	FANCA	IL7	
5	PTCD1	ID3	MMACHC	FANCA	RUSCIAS1	APOE	ZNF670ZNF695	MCM3APAS1	STAG3	IL7	
6	PTCD1	ID3	MMACHC	FANCA	RUSCIAS1	APOE	ZNF670ZNF695	MCM3APAS1	STAG3	IL7	
7	MMACHC	IL7	PTCD1	STAG3	ZNF670ZNF695	APOE	RAETIG	FANCA	SEVERITY	RUSCIAS1	
8	RNU6ATAC	IL7	PTCD1	STAG3	ZNF670ZNF695	APOE	RAETIG	FANCA	SEVERITY	RUSCIAS1	
9	TOMM6	STAG3	SNORA52	THAL	CCDC103	FANCA	APOE4	APOE	MCM3APAS1	MCMD2	
10	RNU6ATAC	IL7	MCMD2	STAG3	ZNF670ZNF695	APOE	SEVERITY	SNORA52	SNORA70	RUSCIAS1	
11	RNU6ATAC	IL7	MCMD2	STAG3	ZNF670ZNF695	APOE	SEVERITY	FANCA	SNORA70	RUSCIAS1	
12	RNU6ATAC	IL7	MCMD2	STAG3	ZNF670ZNF695	APOE	SEVERITY	FANCA	SNORA81	RUSCIAS1	
13	TOMM6	STAG3	SNORA52	THAL	CCDC103	FANCA	APOE4	APOE	MCM3APAS1	MCMD2	
14	MMACHC	RUSCIAS1	PTCD1	STAG3	ZNF670ZNF695	APOE	RAETIG	FANCA	SEVERITY	IL7	
15	MMACHC	RUSCIAS1	PTCD1	STAG3	ZNF670ZNF695	APOE	RAETIG	FANCA	SNORA52	IL7	
16	TOMM6	SNORA52	STAG3	THAL	PTCD1	APOE	APOE4	RUSCIAS1	CCDC103	SNORA81	
17	TOMM6	SNORA52	STAG3	THAL	PTCD1	APOE	APOE4	RUSCIAS1	CCDC103	MCMD2	
18	PTCD1	THAL	SNORA70	CFLARAS1	RNU6ATAC	APOE4	SNORA81	MCM3APAS1	STAG3	IL7	

Table 15 (continued)

Order ranking	9	10	11	12	13	14	15	16	17	18
19	PTCD1	THAL	SNORA70	ID3	RNU6ATAC	APOE4	SNORA81	MCM3APAS1	STAG3	IL7
20	PTCD1	THAL	SNORA70	ID3	RNU6ATAC	APOE4	SNORA81	MCM3APAS1	STAG3	IL7
21	SEVERITY	TOMM6	FANCA	THAL	ID3	PTCD1	SNORA70	APOE	RUSCIAS1	SNORA81
22	TOMM6	SNORA52	STAG3	THAL	PTCD1	APOE	APOE4	RUSCIAS1	SNORA81	CCDC103
23	TOMM6	SNORA52	STAG3	THAL	PTCD1	APOE	APOE4	RUSCIAS1	CCDC103	SNORA81
24	SEVERITY	TOMM6	SNORA70	THAL	ID3	PTCD1	FANCA	APOE	RUSCIAS1	MMACHC
25	SEVERITY	TOMM6	SNORA70	THAL	ID3	PTCD1	FANCA	APOE	RUSCIAS1	MMACHC
26	SEVERITY	TOMM6	FANCA	THAL	ID3	PTCD1	SNORA70	APOE	RUSCIAS1	MMACHC
27	ZNF670ZNF695	TOMM6	STAG3	APOE4	ID1	MCMD2	FANCA	PTCD1	CFLARAS1	THAL
28	TOMM6	STAG3	ID1	THAL	CCDC103	IL7	SNORA81	MMACHC	FANCA	MCMD2
29	SEVERITY	TOMM6	FANCA	STAG3	THAL	PTCD1	SNORA70	APOE	RUSCIAS1	SNORA81
30	RUSCIAS1	MMACHC	THAL	SNORA70	IL7	RNU6ATAC	SNORA52	LOXL3	APOE	APOE4
31	TOMM6	STAG3	ID1	THAL	CCDC103	IL7	SNORA81	MMACHC	APOE	MCMD2
32	LOXL3	MMACHC	THAL	SEVERITY	IL7	RNU6ATAC	SNORA52	RUSCIAS1	APOE	FANCA
33	ZNF670ZNF695	TOMM6	STAG3	APOE4	ID1	CFLARAS1	FANCA	PTCD1	MCMD2	THAL
34	LOXL3	MMACHC	THAL	SNORA70	IL7	RNU6ATAC	SNORA52	RUSCIAS1	APOE	APOE4
35	TOMM6	STAG3	IL7	THAL	CCDC103	ID1	SNORA81	MMACHC	APOE	MCMD2
36	TOMM6	STAG3	SNORA52	THAL	CCDC103	ID1	SNORA81	MMACHC	APOE	MCMD2
37	RUSCIAS1	MMACHC	IL7	SNORA70	THAL	RNU6ATAC	SNORA52	PTCD1	APOE	APOE4
38	RUSCIAS1	MMACHC	IL7	SNORA70	THAL	RNU6ATAC	SNORA52	LOXL3	APOE	APOE4
39	ZNF670ZNF695	TOMM6	STAG3	APOE4	ID1	CFLARAS1	FANCA	PTCD1	MCMD2	THAL
40	APOE	TOMM6	SNORA52	STAG3	THAL	PTCD1	SEVERITY	APOE4	RUSCIAS1	CCDC103
41	RUSCIAS1	MMACHC	CFLARAS1	SNORA70	THAL	RNU6ATAC	SNORA52	SNORA81	MCM3APAS1	APOE4
42	RUSCIAS1	MMACHC	CFLARAS1	SNORA70	THAL	RNU6ATAC	SNORA52	PTCD1	MCM3APAS1	APOE4
43	RUSCIAS1	MMACHC	CFLARAS1	SNORA70	THAL	RNU6ATAC	SNORA52	PTCD1	SNORA81	APOE4
44	LOXL3	MMACHC	RUSCIAS1	PTCD1	STAG3	ZNF670ZNF695	APOE	RAETIG	FANCA	SNORA52
45	LOXL3	MMACHC	RUSCIAS1	PTCD1	STAG3	ZNF670ZNF695	APOE	RAETIG	SNORA52	FANCA
46	SEVERITY	TOMM6	SNORA52	STAG3	THAL	PTCD1	APOE	APOE4	RUSCIAS1	CCDC103
47	SEVERITY	TOMM6	SNORA52	STAG3	THAL	PTCD1	APOE	APOE4	RUSCIAS1	CCDC103
48	ID1	IL7	MCMD2	STAG3	APOE4	APOE	SEVERITY	RNU6ATAC	SNORA70	RUSCIAS1
49	FANCA	IL7	MCMD2	STAG3	APOE4	APOE	SEVERITY	RNU6ATAC	SNORA70	RUSCIAS1
50	ID1	IL7	MCMD2	STAG3	APOE4	APOE	SEVERITY	RNU6ATAC	SNORA70	RUSCIAS1
51	PTCD1	STAG3	SNORA52	THAL	CCDC103	FANCA	APOE4	APOE	MCM3APAS1	MCMD2
52	ID1	IL7	MCMD2	STAG3	ZNF670ZNF695	APOE	SEVERITY	SNORA52	SNORA70	RUSCIAS1
53	ID1	IL7	MCMD2	STAG3	ZNF670ZNF695	APOE	SEVERITY	RNU6ATAC	SNORA70	RUSCIAS1
54	RUSCIAS1	PTCD1	THAL	SNORA70	CFLARAS1	RNU6ATAC	APOE4	SNORA81	MCM3APAS1	STAG3
55	RUSCIAS1	PTCD1	THAL	SNORA70	CFLARAS1	RNU6ATAC	APOE4	SNORA81	MCM3APAS1	STAG3
56	RNU6ATAC	IL7	MCMD2	STAG3	ZNF670ZNF695	APOE	SEVERITY	SNORA52	SNORA70	RUSCIAS1
57	RUSCIAS1	PTCD1	CFLARAS1	SNORA70	THAL	RNU6ATAC	APOE4	SNORA81	MCM3APAS1	STAG3

Table 15 (continued)

Order ranking	19	20	21	22	23	24	25	26	27
1	CCDC103	MMACHC	LOXL3	ID1	NRF1	MCM3APAS1	APOE4	RAET1G	ZNF670ZNF695
2	CFLARAS1	RAET1G	APOE	TOMM6	SNORA81	SNORA52	NRF1	ID1	CCDC103
3	CFLARAS1	MCMD2	APOE4	TOMM6	SNORA81	SNORA52	NRF1	ID1	CCDC103
4	CFLARAS1	MCMD2	APOE4	TOMM6	SNORA81	SNORA52	NRF1	ID1	CCDC103
5	CFLARAS1	RAET1G	APOE4	TOMM6	SNORA81	SNORA52	NRF1	ID1	CCDC103
6	CFLARAS1	MCMD2	APOE4	TOMM6	SNORA81	SNORA52	NRF1	ID1	CCDC103
7	SNORA52	MCMD2	LOXL3	CFLARAS1	SNORA81	TOMM6	NRF1	ID1	CCDC103
8	SNORA52	MCMD2	LOXL3	CFLARAS1	SNORA81	TOMM6	NRF1	ID1	CCDC103
9	PTCD1	SNORA81	ID1	NRF1	RUSCIAS1	SEVERITY	CFLARAS1	ZNF670ZNF695	ID3
10	APOE4	PTCD1	LOXL3	MCM3APAS1	RAET1G	TOMM6	NRF1	ID1	CCDC103
11	APOE4	PTCD1	LOXL3	MCM3APAS1	RAET1G	TOMM6	NRF1	ID1	CCDC103
12	APOE4	PTCD1	LOXL3	MCM3APAS1	RAET1G	TOMM6	NRF1	ID1	CCDC103
13	PTCD1	SNORA81	ID1	NRF1	RUSCIAS1	SEVERITY	LOXL3	ZNF670ZNF695	ID3
14	THAL	MCMD2	APOE4	CFLARAS1	SNORA81	TOMM6	NRF1	ID1	CCDC103
15	THAL	MCMD2	APOE4	CFLARAS1	SNORA81	TOMM6	NRF1	ID1	CCDC103
16	MMACHC	MCMD2	ID1	NRF1	CFLARAS1	IL7	RAET1G	ZNF670ZNF695	ID3
17	MMACHC	SNORA81	ID1	NRF1	CFLARAS1	IL7	RAET1G	ZNF670ZNF695	ID3
18	ID3	RAET1G	APOE	TOMM6	ZNF670ZNF695	SNORA52	NRF1	ID1	CCDC103
19	CFLARAS1	RAET1G	APOE	TOMM6	ZNF670ZNF695	SNORA52	NRF1	ID1	CCDC103
20	CFLARAS1	RAET1G	APOE	TOMM6	ZNF670ZNF695	SNORA52	NRF1	ID1	CCDC103
21	CCDC103	MMACHC	LOXL3	ID1	NRF1	MCM3APAS1	APOE4	RAET1G	ZNF670ZNF695
22	MMACHC	MCMD2	ID1	NRF1	MCM3APAS1	IL7	RAET1G	ZNF670ZNF695	ID3
23	MMACHC	MCMD2	ID1	NRF1	MCM3APAS1	IL7	RAET1G	ZNF670ZNF695	ID3
24	CCDC103	SNORA81	LOXL3	ID1	NRF1	MCM3APAS1	APOE4	RAET1G	ZNF670ZNF695
25	CCDC103	SNORA81	LOXL3	ID1	MCM3APAS1	NRF1	APOE4	RAET1G	ZNF670ZNF695
26	CCDC103	SNORA81	LOXL3	ID1	NRF1	MCM3APAS1	APOE4	RAET1G	ZNF670ZNF695
27	SNORA70	RAET1G	MCM3APAS1	SNORA81	LOXL3	IL7	NRF1	SEVERITY	SNORA52
28	PTCD1	APOE4	MCM3APAS1	NRF1	RUSCIAS1	SEVERITY	LOXL3	ZNF670ZNF695	ID3
29	CCDC103	MMACHC	LOXL3	ID1	NRF1	MCM3APAS1	APOE4	RAET1G	ZNF670ZNF695
30	MCM3APAS1	ID3	STAG3	CFLARAS1	TOMM6	SNORA81	RAET1G	NRF1	ID1
31	PTCD1	APOE4	MCM3APAS1	NRF1	RUSCIAS1	SEVERITY	LOXL3	ZNF670ZNF695	ID3
32	APOE4	ID3	STAG3	CFLARAS1	TOMM6	SNORA81	RAET1G	NRF1	ID1
33	SNORA70	RAET1G	MCM3APAS1	SNORA81	LOXL3	IL7	NRF1	SEVERITY	SNORA52
34	MCM3APAS1	ID3	STAG3	CFLARAS1	TOMM6	SNORA81	RAET1G	NRF1	ID1
35	PTCD1	APOE4	MCM3APAS1	NRF1	RUSCIAS1	SEVERITY	LOXL3	ZNF670ZNF695	ID3
36	PTCD1	APOE4	MCM3APAS1	NRF1	RUSCIAS1	SEVERITY	LOXL3	ZNF670ZNF695	ID3
37	MCM3APAS1	ID3	STAG3	CFLARAS1	TOMM6	SNORA81	RAET1G	NRF1	ID1
38	MCM3APAS1	ID3	STAG3	CFLARAS1	TOMM6	SNORA81	RAET1G	NRF1	ID1
39	SNORA70	RAET1G	RUSCIAS1	SNORA81	LOXL3	IL7	NRF1	SEVERITY	SNORA52

Table 15 (continued)

Order ranking	19	20	21	22	23	24	25	26	27
40	MCMDC2	MMACHC	SNORA81	ID1	NRF1	RAETIG	IL7	CFLARAS1	ZNF670ZNF695
41	PTCD1	ID3	STAG3	APOE	TOMM6	ZNF670ZNF695	RAETIG	NRF1	ID1
42	SNORA81	ID3	STAG3	APOE	TOMM6	ZNF670ZNF695	RAETIG	NRF1	ID1
43	MCM3APAS1	ID3	STAG3	APOE	TOMM6	ZNF670ZNF695	RAETIG	NRF1	ID1
44	IL7	THAL	MCMD2	APOE4	CFLARAS1	SNORA81	TOMM6	NRF1	ID1
45	IL7	THAL	MCMD2	APOE4	CFLARAS1	SNORA81	TOMM6	NRF1	ID1
46	MCMDC2	MMACHC	SNORA81	ID1	NRF1	CFLARAS1	TOMM6	RAETIG	ZNF670ZNF695
47	MCMDC2	MMACHC	SNORA81	ID1	NRF1	RAETIG	IL7	CFLARAS1	ZNF670ZNF695
48	ZNF670ZNF695	PTCD1	LOXL3	MCM3APAS1	RAETIG	TOMM6	NRF1	SNORA52	CCDC103
49	ZNF670ZNF695	PTCD1	LOXL3	MCM3APAS1	RAETIG	TOMM6	NRF1	SNORA52	CCDC103
50	ZNF670ZNF695	PTCD1	LOXL3	MCM3APAS1	RAETIG	TOMM6	NRF1	SNORA52	CCDC103
51	TOMM6	SNORA81	ID1	NRF1	RUSC1AS1	SEVERITY	LOXL3	ZNF670ZNF695	ID3
52	APOE4	PTCD1	LOXL3	MCM3APAS1	RAETIG	TOMM6	NRF1	RNU6ATAC	CCDC103
53	APOE4	PTCD1	LOXL3	MCM3APAS1	RAETIG	TOMM6	NRF1	SNORA52	CCDC103
54	IL7	ID3	RAETIG	APOE	TOMM6	ZNF670ZNF695	SNORA52	NRF1	ID1
55	IL7	ID3	RAETIG	APOE	TOMM6	ZNF670ZNF695	SNORA52	NRF1	ID1
56	APOE4	PTCD1	LOXL3	MCM3APAS1	RAETIG	TOMM6	NRF1	MMACHC	CCDC103
57	IL7	ID3	RAETIG	APOE	TOMM6	ZNF670ZNF695	SNORA52	NRF1	ID1

Methods

Data Access and Characteristics

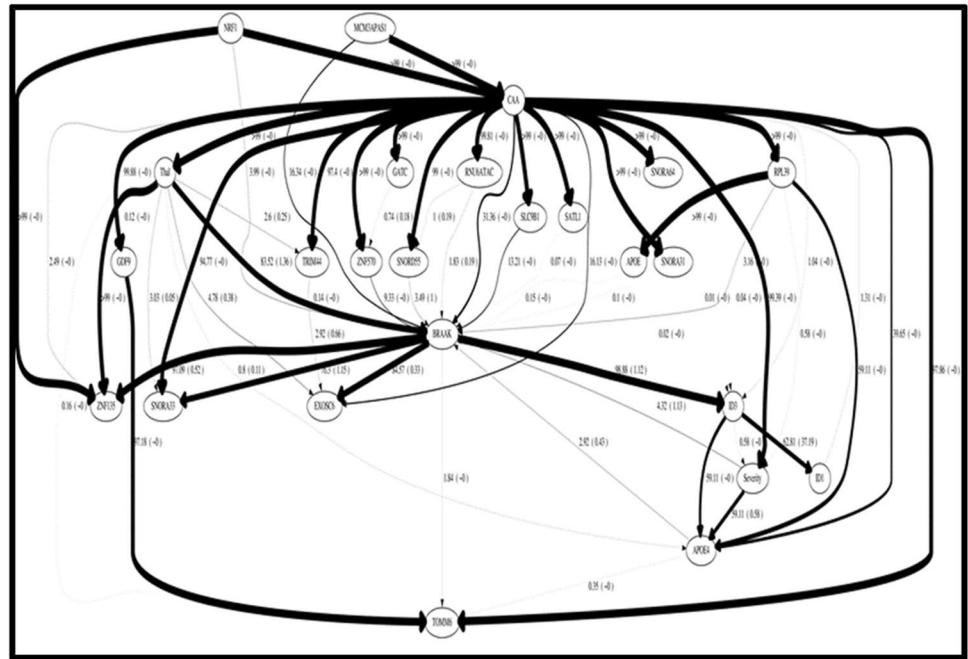
The results published here are in whole or in part based on data obtained from the AD Knowledge Portal. The data available in the AD Knowledge Portal would not be possible without the participation of research volunteers and the contribution of data by collaborating researchers. Data are made available as Open- or Controlled- Access, where individual-level Human data is Controlled Access, and cannot be redistributed.

The RNA-Seq files for CAA cases, AD cases, and controls came from the AD Knowledge Portal. The Mayo Clinic CAA Synapse dataset (syn9779506) was used to access and download CAA cases. The dataset was comprised of two studies derived from the Mayo Clinic Brain Bank. Study 1 of syn9779506, represents Caucasian individuals with primary neuropathological diagnosis of AD and available CAA scores, representing a range of average CAA values between 0 and 4. The collection of genetic and gene expression measures from these individuals was led by Dr. Nilufer Ertekin-Taner at the Mayo Clinic, Jacksonville, FL; the collection of ELISA biochemical measures from these individuals was led by Dr. Guojun Bu. All data was collected as part of the multi-PI RF1AG051504 (MPIs Bu and Ertekin-Taner) using samples from the Mayo Clinic Brain Bank.

Study 2 of syn9779506, represents 75 Caucasian individuals with neuropathological diagnosis of AD, with or without CAA pathology. Of these, 43 had severe CAA pathology (average CAA score ≥ 2 among the five cortical regions). None of the individuals have the following pathologic diagnoses: Parkinson's disease (PD), DLB, PSP, motor neuron disease (MND), CBD, Pick's disease (PiD), Huntington's disease (HD), FTL or dementia lacking distinctive histology (DLDH). The collection of gene expression measures from temporal cortex (TCX) tissue for these individuals was led by Dr. Guojun Bu at the Mayo Clinic, Jacksonville, FL using samples from the Mayo Clinic Brain Bank.

Furthermore, The Mayo Clinic RNA-Seq dataset (syn5550404) was used to download AD Cases and controls for the purposes of this study. This study provides genomic, transcriptomic, and proteomic data from the cerebellum and temporal cortex from individuals from the Mayo Clinic Brain Bank and Banner Sun Health Research Institute. Whole transcriptome data for 275 Cerebellum (CBE) and 276 Temporal cortex (TCX) samples from 312 North American Caucasian subjects with neuropathological diagnosis of AD or elderly controls (CON) without neurodegenerative diseases. Within this cohort, all AD subjects were from the Mayo Clinic Brain Bank (MCBB). Thirty-four control CBE and 31 control TCX samples were

Fig. 19 MCMC ordered NRF1 network for men with CAA



from the MCBB, and the remaining control tissue was from Banner Sun Health Institute. All subjects selected from the MCBB and Banner underwent neuropathologic evaluation by Dr. Dennis Dickson or Dr. Thomas Beach, respectively.

Since CAA is an overlapping pathology in AD patients, we also analyzed RNA-seq data from AD cases. Table 17 summarizes number of CAA cases ($n=275$), sex, age, APOE4 genotype, Braak stage, CAA severity, and brain tissue region (temporal cortex or cerebellum). Table 18 summarizes control subjects described as elderly subjects with a Braak Stage of III or less ($n=150$) and Table 19 summarizes AD cases with a Braak stage of IV or greater ($n=164$).

Gene Expression Analysis

Raw data counts for both CAA and AD cases and controls were aligned and mapped according to the human genome (hg38), implementing STARv2.5.2b [60], and feature Counts from the Subread package v1.5.1 were used for counting of RNA levels [61]. Across cases and controls there were 32,935 common genes and microRNA, which were normalized from raw counts through the trimmed mean of M values (TMM) method in Limma package [62]. We utilized the Limma-Voom approach for differential expression, measuring the precision weights for the mean-variance relationship between CAA/AD cases and controls expression values [62].

For machine learning analysis, we selected candidate genes from CAA cases to be inputted as follows: (i) genes with ± 2 FC across CAA cases vs controls, (ii) genes found to be significant (p -value ≤ 0.05 , student t-test), and (iii)

genes that are ID3 or NRF1 targets. Using this criterion, we created four cohorts to input into machine learning analysis based on sex and ID3/NRF1 significant target genes.

We transformed the normalized RNA Count levels into z-scores per gene and discretized the z-values into the categories less than -1 ($z < -1$), between -1 and 1 ($-1 \leq z \leq 1$), and more than 1 ($z > 1$) to represent low expression, no change in expression, and high expression of a gene, respectively [63]. The clinical criteria for CAA was codified as follows: (i) CAA (control (0), case (1)), (ii) Thal phase (≤ 1 (0), ≥ 1 to 4 (1), > 4 (2)), (iii) BRAAK Score (< 5 (1), > 5 (2)), (iv) APOE4 carrier (APOE4 Non-Carrier (1), APOE4 Carrier (2)), (v) CAA Severity (0 to 1.9 (0), 2 to 2.5 (1), 2.5 or greater (2)), (vi) Age (≤ 79 years old (1), > 79 years old (2)).

Microarray Analysis

We further validated whether high ID3 and low NRF1 mRNA expression occurred in brain microvessels from AD patients by downloading microarray data file GSE45596 [64] from the NCBI Gene Expression Omnibus (GEO). Raw intensity values for the dataset were normalized according to their log2 ratios and then analyzed using limma differential expression analysis.

Transcription Regulating Protein Activity

To determine if high ID3 and low NRF1 mRNA levels observed in CAA patients impacted genome wide transcriptional activity, we conducted a transcriptional activity assay

Table 16 Markov blanket MCMC causal gene orders of CAA male cases/controls and NRF1 targets

Order ranking	Order probability	1	2	3	4	5	6	7	8	9	10	11	12
1	0.5211569	MCM3AP-AS1	NRF1	CAA	Thal	SNORA31	GATC	RPL39	RNU6ATAC	TRIM44	SLC9B1	APOE	ZNF570
2	0.1550542	MCM3AP-AS1	NRF1	CAA	Thal	SNORD55	GATC	RPL39	TRIM44	RNU6ATAC	SLC9B1	APOE	ZNF570
3	0.1940341	MCM3AP-AS1	NRF1	CAA	Thal	SNORD55	GATC	RPL39	RNU6ATAC	TRIM44	SLC9B1	APOE	ZNF570
4	0.0419495	MCM3AP-AS1	NRF1	CAA	Thal	SNORD55	EXOSC6	RPL39	RNU6ATAC	TRIM44	SLC9B1	APOE	ZNF570
5	0.0080037	ID1	MCM3AP-AS1	NRF1	CAA	Thal	GATC	SNORA31	RPL39	APOE4	TRIM44	SLC9B1	APOE
6	0.0073252	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA64	SNORA31	RPL39	APOE4	TRIM44	SLC9B1	APOE
7	0.0049023	ID1	MCM3AP-AS1	NRF1	CAA	Thal	GATC	SNORA31	RPL39	APOE4	TRIM44	SLC9B1	APOE
8	0.0081694	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA31	GATC	RPL39	APOE4	TRIM44	SLC9B1	APOE
9	0.0170413	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA64	SNORA31	RPL39	APOE4	TRIM44	SLC9B1	APOE
10	0.0060874	MCM3AP-AS1	NRF1	CAA	Thal	SNORA31	GATC	RPL39	APOE4	TRIM44	SLC9B1	APOE	ZNF570
11	0.001298	MCM3AP-AS1	NRF1	CAA	SNORA64	Severity	GATC	RPL39	SNORA31	RNU6ATAC	BRAAK	APOE	EXOSC6
12	0.0015255	MCM3AP-AS1	NRF1	CAA	SNORA64	Severity	GATC	RPL39	SNORA31	RNU6ATAC	BRAAK	APOE	EXOSC6
13	0.0010907	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA33	SNORA31	RPL39	APOE4	APOE	Severity	SNORD55
14	0.0051401	MCM3AP-AS1	NRF1	CAA	SNORA64	Severity	GATC	RPL39	SNORA31	RNU6ATAC	BRAAK	APOE	EXOSC6
15	0.0013596	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA33	SNORA31	RPL39	APOE4	TRIM44	Severity	SNORD55
16	0.0040324	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA33	SNORA31	RPL39	APOE4	APOE	Severity	SNORD55
17	0.0005137	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA33	SNORA31	RPL39	APOE4	TRIM44	SNORA64	SNORD55
18	0.0013938	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA64	SNORA31	RPL39	APOE4	TRIM44	SNORA33	SNORD55
19	0.0023632	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA64	SNORA31	RPL39	APOE4	TRIM44	SNORA33	APOE
20	0.0003046	MCM3AP-AS1	NRF1	CAA	SNORA64	Severity	GATC	RPL39	SNORA31	RNU6ATAC	BRAAK	APOE	EXOSC6
21	0.0017362	MCM3AP-AS1	NRF1	CAA	SNORA64	Severity	GATC	RPL39	SNORA31	RNU6ATAC	BRAAK	APOE	EXOSC6
22	0.0042831	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA33	SNORA31	RPL39	APOE4	APOE	BRAAK	SNORD55
23	0.0040362	MCM3AP-AS1	NRF1	CAA	Thal	SNORD55	GATC	RPL39	RNU6ATAC	TRIM44	SLC9B1	APOE	ZNF570
24	0	ID1	MCM3AP-AS1	NRF1	CAA	Thal	SNORA33	SNORA31	RPL39	APOE4	TRIM44	ZNF135	SNORD55
Order ranking	13	14	15	16	17	18	19	20	21	22	23	24	25
1	SATL1	BRAAK	ID3	Severity	GDF9	APOE4	SNORA64	SNORD55	ZNF135	SNORA33	TOMM6	EXOSC6	ID1
2	SATL1	BRAAK	ID3	Severity	GDF9	APOE4	SNORA64	SNORA31	ZNF135	SNORA33	TOMM6	EXOSC6	ID1
3	SATL1	BRAAK	ID3	Severity	GDF9	APOE4	SNORA64	SNORA31	ZNF135	SNORA33	TOMM6	EXOSC6	ID1
4	SATL1	BRAAK	ID3	Severity	GDF9	APOE4	SNORA64	SNORA31	ZNF135	SNORA33	TOMM6	GDF9	ID1
5	ZNF570	SATL1	BRAAK	ID3	Severity	GDF9	RNU-6ATAC	SNORA64	SNORA33	ZNF135	SNORD55	TOMM6	EXOSC6
6	ZNF570	SATL1	BRAAK	ID3	Severity	GDF9	RNU-6ATAC	GDF9	SNORA33	ZNF135	SNORD55	TOMM6	EXOSC6
7	ZNF570	SATL1	BRAAK	ID3	Severity	GDF9	RNU-6ATAC	SNORA64	SNORD55	ZNF135	SNORA33	TOMM6	EXOSC6
8	ZNF570	SATL1	BRAAK	ID3	Severity	GDF9	RNU-6ATAC	SNORA64	SNORD55	ZNF135	SNORA33	TOMM6	EXOSC6

Table 16 (continued)

Order ranking	13	14	15	16	17	18	19	20	21	22	23	24	25
9	ZNF570	SATL1	BRAAK	ID3	Severity	GDF9	SNORD55	GDF9	SNORA33	ZNF135	RNU-6ATAC	TOMM6	EXOSC6
10	SATL1	BRAAK	ID3	Severity	GDF9	RNU-6ATAC	SNORA64	SNORD55	ZNF135	SNORA33	TOMM6	EXOSC6	ID1
11	Thal	TRIM44	ID3	APOE4	SNORA33	SATL1	SNORD55	SLC9B1	GDF9	ZNF135	TOMM6	ID1	ZNF570
12	Thal	TRIM44	ID3	APOE4	GDF9	SATL1	SNORD55	SLC9B1	SNORA33	ZNF135	TOMM6	ID1	ZNF570
13	ZNF570	SLC9B1	BRAAK	ID3	SNORA64	GDF9	TRIM44	GDF9	SATL1	ZNF135	RNU-6ATAC	TOMM6	EXOSC6
14	Thal	TRIM44	ID3	SATL1	GDF9	APOE4	SNORD55	SLC9B1	SNORA33	ZNF135	TOMM6	ID1	ZNF570
15	ZNF570	SATL1	BRAAK	ID3	SNORA64	GDF9	APOE	GDF9	SLC9B1	ZNF135	RNU-6ATAC	TOMM6	EXOSC6
16	ZNF570	SATL1	BRAAK	ID3	SNORA64	GDF9	TRIM44	GDF9	SLC9B1	ZNF135	RNU-6ATAC	TOMM6	EXOSC6
17	ZNF570	SATL1	BRAAK	ID3	Severity	GDF9	APOE	GDF9	SLC9B1	ZNF135	RNU-6ATAC	TOMM6	EXOSC6
18	ZNF570	SATL1	BRAAK	ID3	Severity	GDF9	APOE	GDF9	SLC9B1	ZNF135	RNU-6ATAC	TOMM6	EXOSC6
19	ZNF570	SATL1	BRAAK	ID3	Severity	GDF9	SNORD55	GDF9	SLC9B1	ZNF135	RNU-6ATAC	TOMM6	EXOSC6
20	SNORD55	ID3	TRIM44	APOE4	SNORA33	SATL1	Thal	SLC9B1	GDF9	ZNF135	TOMM6	ID1	ZNF570
21	SNORD55	ID3	TRIM44	APOE4	SNORA33	SATL1	Thal	SLC9B1	GDF9	ID1	TOMM6	ZNF135	ZNF570
22	ZNF570	SLC9B1	Severity	ID3	SNORA64	GDF9	TRIM44	GDF9	SATL1	ZNF135	RNU-6ATAC	TOMM6	EXOSC6
23	SATL1	SNORA31	ID3	Severity	GDF9	APOE4	SNORA64	BRAAK	ZNF135	SNORA33	TOMM6	EXOSC6	ID1
24	ZNF570	SATL1	BRAAK	ID3	Severity	GDF9	APOE	GDF9	SLC9B1	SNORA64	RNU-6ATAC	TOMM6	EXOSC6

using LRPATH logistic regression analysis [65]. Given a high-throughput dataset with continuous significance values (i.e. p-values), LRpath tests for gene sets (termed concepts) with significantly higher significance values (e.g. for differential expression) than expected at random. Using LRPath we examined the total transcriptional activity of ID3 and NRF1 based on the expression of their target genes for discovery of CAA-associated and AD-associated transcriptional activity. The Encyclopedia of DNA Elements (ENCODE) was used to determine the ChIP-seq targets for NRF1 and ID3 to conduct transcriptional activity analysis. Transcriptional activity of ID3 and NRF1 was determined from the cumulative statistically significant differential expression of hundreds of 9483 ID3 target genes and 11,280 NRF1 target genes respectively. A total of 9438 ID3 gene targets from ENCODE were used to determine its total transcriptional regulating activity. A total of 11,280 NRF1 bound target genes from ENCODE were used to determine total transcription factor activity based of both p-value and logFC values across CAA and AD cases in both sexes.

ID3 and NRF1 genome wide transcriptional activity was also determined by pathways including VEGF signaling, apoptosis, cell cycle, cell senescence, and angiogenesis. The human targets in the VEGF Signaling, Apoptosis, Cell Cycle, Cellular Senescence, and Epigenetic Chromatin Remodeling Factor pathways were determined by using SA Biosciences (SAB) human target gene lists. Furthermore, for unique human endothelial cell signatures we cross compared endothelial pathway targets derived from global expression analysis of human microvascular endothelial cells undergoing biomechanical stress [66]. For the brain endothelial signature, we determined target genes derived from single cell RNA-Seq analysis of brain endothelial cells derived from *Mus Musculus* and *C57BL/6 J* mice [67, 68]. We then cross-compared targets found in each of the lists with 9438 ID3 ChIP-seq target genes to create a custom list of only ID3 target genes in each of the signature pathways to determine ID3 driven transcriptional activity and did the same procedure with the 11,280 NRF1 genes derived from ENCODE ChIP-seq targets to determine NRF1 driven transcriptional activity.

Causal Bayesian Network Modeling: Structural and Parameter Learning

In clinical and research fields, machine learning methods are utilized to study statistical relationships in disease progression through the creation of causal networks from high throughput transcriptomic datasets [69, 70]. Statistical machine learning methods help to identify key upstream regulators from a causal network inferred from RNA-Seq genomic data and clinical data. Causal Bayesian Networks

(CBNs) are used as means to learn the causal networks inferred from genomic data. A CBN consists of directed acyclic graphs (DAGs) comprised of nodes, that represent the random variables (genes and clinical variables) being modeled, and intervening arcs (arrows), which represent the relationships between these random variables [71]. The nodes of the networks represent the expression of the genes and clinical variables. Resulting networks represent a graphic representation of the causal hypothesis [72].

To learn the causal Bayesian networks for our study, we used Bayesian Network Inference with Java Objects (BANJO), which is a computational modeling tool based on a data-driven method using Bayesian network frameworks to obtain directed inference networks [69, 71]. The purpose of BANJO was to identify gene-gene interactions given our datasets of CAA cases and controls to better understand the clinical/pathological characteristics of the disease when ID3 and NRF1 act as transcriptional drivers of these networks. The scoring metric used is called Bayesian Dirichlet equivalence (BDe). BANJO keeps making incremental changes in the structure to improve the BDe score of the network [73]. The final DAG shows regulation between genes and their possible involvement in the outcome (disease). The goal of this analysis was to create a proposed network showing the interaction among these variables in CAA to identify possible drivers for the disease. Computational learning of causal Bayesian networks takes time and computational power, therefore, to have parsimonious models all Bayesian networks learned from the data were limited to the possible number of parents of five across all four cohorts (ID3 Female, ID3 Male, NRF1 Female, NRF1 Male).

The first-degree Markov blanket (MB) of a variable X in a CBN (denote as MB [X]) is defined as the set of variables consisting of disease causes (parents) of X and direct effects (children) of X and all other direct causes (parents) of the direct effects (children) of X [20]. We calculated the first-degree Markov blanket for variable CAA in the CBN (denoted as Markov blanket [CAA]) across the four cohorts was defined as the set of variables that represents the direct causes (parents) of the direct effects (children) of CAA.

ID3 Female For causal Bayesian network structure learning, we performed independent runs in BANJO across four separate timepoints, i.e., 2 Hr, 6 Hr, 12 Hr, and 24 h. At each timepoint, we performed three independent structure learning networks. Therefore, a total of 12 runs of independent causal Bayesian network structure learning with total of $3 \times 2 \text{ h} + 3 \times 6 \text{ h} + 3 \times 12 \text{ h} + 3 \times 24 \text{ h} = 132 \text{ h}$ of runs were performed. Using the dataset for ID3 Female CAA cases vs controls, we outputted the 12 best log-likelihood structures reported for each run, and we selected the network with the highest log-likelihood structure at each timepoint as the initial structure for subsequent timepoints. Once

improvement of the model stopped at subsequent timepoints we stopped running the model. We then compared the first-degree Markov blanket genes for the CAA node across the top 3 scoring networks in our trial runs and selected the common genes for subsequent validation runs. Following, Markov blanket gene comparison for CAA the original CBN

structure went from 318 to 74 variables for further validation runs. The first MB genes for CAA were then used for MCMC gene ordering and optimal gene structure learning.

ID3 Male For CBN structure learning, we performed independent runs across four separate timepoints, i.e., 2 Hr, 6

Table 17 Demographic Table of CAA cases from The Mayo Clinic CAA-AD study (syn9779506)

Sex M = Males F = Females	Age at death	Age spans	ApoE genotype	Braak stage	CAA severity	Brain region (Cer = Cerebellum, TC = Temporal Cortex)
M = 129	$\mu = 79.30$ yrs $\alpha = 7.84$ yrs Max = 90 or above yrs Min = 55 yrs	≤ 90 years old = 8 80 to 89 yrs. old = 66 70 to 79 yrs. old = 42 60 to 69 yrs. old = 8 < 60 yrs old = 5	ApoE 2 = 2/3 = 0 ApoE 2 = 2/4 = 4 ApoE 3 = 3/3 = 43 ApoE 3 or 4 = 3/4 = 60 ApoE 4 = 4/4 = 22	$(\geq 6) = 49$ (5–5.9) = 55 (4 to 4.9) = 25 (0 – 3.9) = 0	$(> 4) = 0$ (3–3.9) = 4 (2–2.9) = 30 (1–1.9) = 27 (0 – 0.9) = 87	Cer = 89 TC = 40
F = 146	$\mu = 82.794$ yrs $\alpha = 7.32$ yrs Max = 90 or above yrs Min = 55 yrs	≤ 90 years old = 33 80 to 89 yrs. old = 76 70 to 79 yrs. old = 30 60 to 69 yrs. old = 3 < 60 yrs old = 4	ApoE 2 = 2/3 = 0 ApoE 2 = 2/4 = 2 ApoE 3 = 3/3 = 40 ApoE 3 or 4 = 3/4 = 79 ApoE 4 = 4/4 = 25	$(\geq 6) = 76$ (5–5.9) = 54 (4 to 4.9) = 16 (0 – 3.9) = 0	$(> 4) = 1$ (3–3.9) = 2 (2–2.9) = 26 (1–1.9) = 30 (0 – 0.9) = 87	Cer = 111 TC = 35

Table 18 Demographic table of controls from the Mayo RNA-seq study (syn5550404)

Sex M = Males F = Females	Age at death	Age spans	ApoE genotype	CAA severity	Brain region (Cer = Cer- ebellum, TC = Temporal Cortex)
M = 82	$\mu = 82.50$ yrs $\alpha = 8.06$ yrs Max = 90 or above yrs Min = 58 yrs	≤ 90 years old = 21 80 to 89 yrs. old = 39 70 to 79 yrs. old = 14 60 to 69 yrs. old = 5 < 60 yrs old = 2	ApoE 2 = 2/3 = 12 ApoE 2 = 2/4 = 1 ApoE 3 = 3/3 = 59 ApoE 3 or 4 = 3/4 = 9 ApoE 4 = 4/4 = 0	$(> 4) = 0$ (3–3.9) = 0 (2–2.9) = 0 (1–1.9) = 0 (0 – 0.9) = 81	Cer = 41 TC = 41
F = 73	$\mu = 82.01$ yrs $\alpha = 8.86$ yrs Max = 90 or above yrs Min = 53 yrs	≤ 90 years old = 16 80 to 89 yrs. old = 36 70 to 79 yrs. old = 14 60 to 69 yrs. old = 4 < 60 yrs old = 3	ApoE 2 = 2/3 = 10 ApoE 2 = 2/4 = 1 ApoE 3 = 3/3 = 54 ApoE 3 or 4 = 3/4 = 8 ApoE 4 = 4/4 = 0	$(> 4) = 0$ (3–3.9) = 0 (2–2.9) = 0 (1–1.9) = 0 (0 – 0.9) = 73	Cer = 36 TC = 37

Table 19 Demographic table of AD cases from the Mayo RNA-seq study (syn5550404)

Sex M = Males F = Females	Age at death	Age spans	ApoE genotype	Braak stage	Brain region (Cer = Cer- ebellum, TC = Temporal Cortex)
M = 67	$\mu = 79.50$ yrs $\alpha = 7.73$ yrs Max = 90 or above yrs Min = 63 yrs	≤ 90 years old = 14 80 to 89 yrs. old = 31 70 to 79 yrs. old = 14 60 to 69 yrs. old = 8	ApoE 2 = 2/3 = 3 ApoE 2 = 2/4 = 0 ApoE 3 = 3/3 = 35 ApoE 3 or 4 = 3/4 = 26 ApoE 4 = 4/4 = 3	$(\geq 6) = 29$ (5–5.9) = 32 (4 to 4.9) = 6 (0 – 3.9) = 0	Cer = 34 TC = 33
F = 97	$\mu = 80.8$ yrs $\alpha = 6.90$ yrs Max = 90 or above yrs Min = 60 yrs	≤ 90 years old = 26 80 to 89 yrs. old = 47 70 to 79 yrs. old = 18 60 to 69 yrs. old = 6	ApoE 2 = 2/3 = 5 ApoE 2 = 2/4 = 0 ApoE 3 = 3/3 = 35 ApoE 3 or 4 = 3/4 = 48 ApoE 4 = 4/4 = 9	$(\geq 6) = 26$ (5–5.9) = 47 (4 to 4.9) = 18 (0 – 3.9) = 6	Cer = 48 TC = 49

Hr, 12 Hr, and 24 h. At each timepoint, we performed three independent CBN structure learning with total of $3 \times 2 \text{ h} + 3 \times 6 \text{ h} + 3 \times 12 \text{ h} + 3 \times 24 \text{ h} = 132 \text{ h}$ of runs were performed. Using the dataset for ID3 male CAA cases vs controls, we outputted the 12 best log-likelihood structures reported for each run, and we selected the network with the highest log-likelihood structure at each timepoint as the initial structure for subsequent timepoints. Once improvement of the model stopped at subsequent timepoint we stopped running the model. The first-degree Markov blanket of variable CAA in the CBN (denoted as Markov blanket [CAA]) was defined as the set of variables that represents the direct causes (parents) of the direct effects (children) of CAA. We then compared the first-degree Markov blanket genes for the CAA node across the top 3 scoring networks in our trials runs and selected the common genes for subsequent validation runs. Following, Markov blanket gene comparison for CAA the original CBN structure went from 270 to 97 variables for further validation runs. The first MB genes for CAA were then used for MCMC gene ordering and optimal gene structure learning.

NRF1 Female For CBN structure learning, we performed independent runs across three separate timepoints, i.e., 2 Hr, 6 Hr, and 12 Hr. At each timepoint, we performed three independent CBN structure learning with total of $3 \times 2 \text{ h} + 3 \times 6 \text{ h} + 3 \times 12 \text{ h} = 60 \text{ h}$ of runs until the network had improved until the network reached its best log-likelihood CBN. Using the dataset for NRF1 female CAA cases vs controls, we outputted the 9 best log-likelihood structures reported for each run, and we selected the network with the highest log-likelihood structure at each timepoint as the initial structure for subsequent timepoints. Improvement was measured according to log normalization of Bayesian Dirichlet (BDE) scores at each timepoint the model was run. The first-degree Markov blanket of variable CAA in the CBN (denoted as Markov blanket [CAA]) was defined as the set of variables that represents the direct causes (parents) of the direct effects (children) of CAA. Following, Markov blanket gene comparison for CAA the original CBN structure went from 312 to 102 variables for further validation runs. The first MB genes for CAA were then used for MCMC gene ordering and optimal gene structure learning.

NRF1 Male For CBN structure learning, we also performed independent runs across three separate timepoints, i.e., 2 Hr, 6 Hr, and 12 Hr. At each timepoint, we performed three independent CBN structure learning networks for a total of $3 \times 2 \text{ h} + 3 \times 6 \text{ h} + 3 \times 12 \text{ h} = 60 \text{ h}$ of runs until the network had improved until the network reached its best log-likelihood CBN. We used the dataset for NRF1 male CAA cases vs controls, we outputted the 9 best log-likelihood structures reported for each run, and we selected the network with the

highest log-likelihood structure at each respective timepoint for the initial structure of subsequent timepoints. The first-degree Markov blanket of variable CAA in the CBN was defined as the set of variables that represents the direct causes (parents) of the direct effects (children) of CAA. In order to determine genes for our validation runs, we conducted a Markov blanket gene comparison for the CAA node across the three best scoring networks, and this led to a determination of 94 variables to run in our validation network. The first MB genes for CAA were then used for MCMC gene ordering and optimal gene structure learning.

CBN Validation of Learned Networks and Sensitivity Analysis of CAA Probability in GeNIe

Based on the optimal scoring (BDe score) proposed probabilistic graphical model structures for ID3 and NRF1 target genes across both males and females, we learned the probable contribution of 1st degree Markov Blanket target genes to CAA. The best scoring probabilistic graphical model for ID3 consisted of 21 variables including CAA, meanwhile the best scoring probabilistic graphical model for NRF1 consisted of 22 variables including CAA. To calculate the probabilities of the variables from the proposed probabilistic graphic model and to test the sensitivity of CAA status to changes in the gene expression of Markov genes, we recreated the BANJO network structure using the GeNIe modeler [74]. Learning the parameters of the Markov blanket genes for the CAA node and its connected variables was performed with GeNIe, a Bayesian network graphical interface tool developed by BayesFusion [74]. After the structure and parameters for each node were assembled in GeNIe, a sensitivity analysis was done by modifying the evidence (marginal probability) of different nodes (gene expression stages) thereby observing the effect on the probability of the CAA node. We estimated the conditional and marginal probability distributions of CAA by changing the expression levels (evidence) of one or more ID3/NRF1 target genes. We then modeled disease risk by changing expression levels of single genes (ID3, NRF1, etc.) or groups of genes denoted as $P(\text{CAA} \mid \text{Modified Markov blanket gene expression level})$.

Estimation of Lifetime CAA Relative Risk (RR)

We used the 100% marginal probability of high, medium, and low levels of gene expression of each individual Markov blanket gene as well as combinations of two or more genes and estimated the percent probability of CAA [Pr. (CAA)] for each case. To account for lifetime risk of CAA amongst these first-degree MB genes in both ID3 and NRF1 optimal structures we

needed to determine $P(\text{CAA})$ in the population. Further validation of the effects of Markov blanket genes on risk of CAA was determined in both the ID3 and NRF1 optimal structure by using Bayes' theorem:

$$\begin{aligned} P(\text{CAA}|G_1, G_2, \dots, G_n) \\ = P(G_1, G_2, \dots, G_n|\text{CAA})P(\text{CAA}) \\ /P(G_1, G_2, \dots, G_n) \end{aligned}$$

The prevalence of CAA in the general population ages 79–84 yrs old has been reported to be 8% [75]. Therefore, we calculated $P(\text{CAA})=0.08$ as baseline risk for the general adult population and used this in the lifetime risk calculations.

Estimation of the Minimum set of Combination of Gene Expression Patterns for Maximum Relative Risk (RR) using PREDICT Combinatorial Analysis

Genie's 'learn parameters' function analysis of the 16 genes (including ID3) associated with the ID3 representative network showed that 5 MB genes (ID3, ATG10, SLC5A2, U2AF1, and UTS2) consistently distinguished between control and CAA cases. Furthermore, the NRF1 representative network Genie's 'learn' parameters calculated the probability of evidence for CAA likelihood given the modified expression levels of the 18 genes (including NRF1 and ID3) associated with the causal Bayesian network for CAA in men. Consistently, 5 MB gene targets (GATC, GDF9, TOMM6, SLC9B1, ZNF135) of NRF1 were found to contribute most to the likelihood of CAA in the PREDICT combinatorial analysis.

Amongst the ID3 representative network in females and NRF1 representative network in males, MB genes were calculated by different gene configurations g with the collected dataset, and we used the SMILE library in Bayes Fusion and the C++ program (PREDICT) to calculate the following:

$$P(D|R = g),$$

where D represents a subject has CAA and $R = \{\text{ATG10, ID3, SLC5A2, U2AF1, UTS2}\}$ for the ID3 representative network and $R = \{\text{GATC, GDF9, NRF1, SLC9B1, TOMM6, ZNF135}\}$ for the NRF1 representative network. Among the gene configurations g that predicts CAA with high or low probability (i.e., $P(D|R = g) > 0.99999$ or $P(D|R = g) < 1.0 \times 10^{-6}$), we focused on g where ID3 and NRF1 were either expressed high or low.

To find the minimum set of combination of gene expression patterns that provide us with the maximum relative risk (RR), we calculated the following using PREDICT program:

$$S = \arg \max \frac{P(D|Q = q)}{P(D|Q = q')}$$

where Q represents any subset of R ; S represents a set of the minimum number of genes that maximizes the RR term, $q = \arg \max_g P(D|Q = g)$ and $q' = \arg \min_g P(D|Q = g)$ with q and q' representing two different gene expression patterns among the genes in S that maximizes and minimizes $P(D|Q)$, respectively [73].

For the 5 ID3 representative network distinguished genes (including ID3), we report the 7 maximum RR ratios of the minimum set of combination of gene expression patterns generated by PREDICT combinatorial analysis. Furthermore, for the 6 NRF1 representative network distinguished genes (including NRF1), we report the 7 maximum RR ratios of the minimum set of combination of gene expression patterns generated by PREDICT combinatorial analysis.

MCMC Ordering Analysis

To further illustrate sex-specific higher order regulatory relationships between methodologically identified ID3 targets and NRF1 motif-enriched genes, CAA disease state, and relevant clinical variables, we employed Monte Carlo over Markov Chain (MCMC) simulation and optimal structure learning algorithms established in SMILE (Structural Modeling, Inference, and Learning Engine), which is a learning/causal discovery engine for graphical models, such as Bayesian networks, influence diagrams, and structural equation models available from BayesFusion, LLC. SMILE is an organized library of C++ classes that can be overlaid into existing user software through its API, thereby augmenting user products with several decision-making capabilities. Previously, we have ran MCMC order search as described across several studies [18]. The MCMC ordering results were then followed by a graphical output visualization through the R package bnlearn displaying weighted arcs. Across sex-specific networks, ID3 and NRF1 targeted genes inputted into the MCMC ordering algorithm were candidates that were identified as first-degree MB gene candidates and possessed influence scores above 0.70 as determined by the BANJO algorithm. Furthermore, clinical variables (THAL phase, CAA severity, age, APOE4 carrier status, BRAAK) unique to patient cohort were inputted as part of MCMC ordering search as well. Across sex-specific networks causal probabilities were calculated between arcs (connections) in the CBNs for ID3 and NRF1 generated from the BANJO algorithm to help determine representative structures derived from the MCMC ordering analysis. Here we detail all ordering networks within > 99% probable orders, and for each order, we report a subsequent representative structure that is among the > 99% probable structures. In each of the graphical networks reported from the MCMC ordering method, > 99% represents a probability percentage

of 99.99 and above, ~ 0 represents a probability percentage of 0.01 and below, and percentages in the parenthesis of arcs (connections) in the representative structure is the probability that relationship between the arcs can be reversed in direction.

Acknowledgements This study was possible because of the participation of research volunteers and the contribution of data by collaborating researchers to the AD Knowledge Portal. This study was supported by funds from the Robert Stempel College of Public Health and Social Work, the National Science Foundation CREST Award (#1547798) and the National Institutes of Health R15 Award (1R15HL145652-01).

Authors' Contributions Christian Perez performed the machine learning and causal gene order experiments as well as performed the RNA-seq analysis. He also contributed in the writing and editing of the manuscript. Zhenghua Gong contributed to the methodology of machine learning, combinatorial risk, and MCMC causal gene ordering. Changwon Yoo designed the methodology for machine learning, combinatorial risk, and causal gene ordering experiments; and contributed to the writing, reviewing, and editing of the manuscript. Deodutta Roy and Alok Deoraj contributed to the conceptualization, writing, and editing of the manuscript. Quentin Felty provided funding, administrated the project, and contributed to the conceptualization, writing, and editing of the manuscript.

Funding This work was partially supported by the National Science Foundation CREST Award (#1547798 and the National Institutes of Health R15 Award (1R15HL145652-01).

Data Availability The results published here are in whole or in part based on data obtained from the AD Knowledge Portal. Details of analysis used in the current study are available from the first author or corresponding author upon request. Data are made available as open- or controlled-access. The RNA-Seq datasets analyzed throughout the study are available in the AD Knowledge Portal Synapse Repository, at syn9779506 and syn5550404.

The Mayo Clinic AD-CAA study was led by Dr. Guojun Bu and Dr. Nilufer Ertekin-Taner at the, Mayo Clinic, Jacksonville, FL as part of the multi-PI RF1AG051504 (MPIs Bu and Ertekin-Taner) using samples from the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50AG016574, R37AG027924, Cure Alzheimer's Fund, and support from Mayo Foundation. The Mayo RNAseq study data was led by Dr. Nilüfer Ertekin-Taner, Mayo Clinic, Jacksonville, FL as part of the multi-PI U01 AG046139 (MPIs Golde, Ertekin-Taner, Younkin, Price). Samples were provided from the following sources: The Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona.

The microvessel microarray dataset analyzed in this study are available in the Gene Expression Omnibus at the National Center for Biotechnology Information at GSE45596.

Declarations

Ethics Approval This is a secondary data analysis study. The FIU Research Ethics Committee confirmed that no ethical approval was required and granted the IRB-20-0304 exemption to do a secondary data analysis of the RNA-seq obtained from the AD Knowledge Portal.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare that they have no competing interests.

References

1. Biffi A, Greenberg SM (2011) Cerebral amyloid angiopathy: a systematic review. *J Clin Neurol* [Internet] 7(1):1. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3079153/>
2. Viswanathan A, Greenberg SM (2009) Intracerebral hemorrhage. *Handb Clin Neurol* 93:767–790. [https://doi.org/10.1016/S0072-9752\(08\)93038-4](https://doi.org/10.1016/S0072-9752(08)93038-4)
3. van Asch CJ, Luitse MJ, Rinkel GJ, van der Tweel I, Algra A, Klijn CJ (2010) Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. *Lancet Neurol*. [cited 2022 Jun 2];9(2):167–76. Available from: <https://pubmed.ncbi.nlm.nih.gov/20056489/>
4. Qureshi AI, Tuhrim S, Broderick JP, Batjer HH, Hondo H, Hanley DF (2001) Spontaneous intracerebral hemorrhage. *N Engl J Med*. [cited 2022 Jun 2];344(19):1450–60. Available from: <https://pubmed.ncbi.nlm.nih.gov/11346811/>
5. Pfeifer LA, White LR, Ross GW, Petrovitch H, Launer LJ (2002) Cerebral amyloid angiopathy and cognitive function. *Neurology*. [cited 2023 Feb 18];58(11):1629–34. Available from: <https://n.neurology.org/content/58/11/1629>
6. Verghese PB, Castellano JM, Garai K, Wang Y, Jiang H, Shah A, Bu G, Frieden C, et al (2013) ApoE influences amyloid- β ($A\beta$) clearance despite minimal apoE/ $A\beta$ association in physiological conditions. *Proc Natl Acad Sci USA*. [cited 2023 Feb 18];110(19):E1807–16. Available from: <https://www.pnas.org/doi/abs/10.1073/pnas.1220484110>
7. Nelson PT, Pious NM, Jicha GA, Wilcock DM, Fardo DW, Estus S, Rebeck GW (2013) APOE- ϵ 2 and APOE- ϵ 4 correlate with increased amyloid accumulation in cerebral vasculature. *J Neuropathol Exp Neurol* 72(7):708–715. <https://doi.org/10.1097/NEN.0b013e31829a25b9>
8. Altmann A, Tian L, Henderson VW, Greicius MD (2014) Sex modifies the APOE-related risk of developing Alzheimer disease. *Ann Neurol*. [cited 2023 Feb 18];75(4):563–73. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/ana.24135>
9. Shinohara M, Murray ME, Frank RD, Shinohara M, DeTure M, Yamazaki Y, Tachibana M, Atagi Y, et al (2016) Impact of sex and APOE4 on cerebral amyloid angiopathy in Alzheimer's disease. *Acta Neuropathol*. [cited 2023 Feb 18];132(2):225–34. Available from: <https://link.springer.com/article/10.1007/s00401-016-1580-y>
10. Perez CM, Felty Q (2022) Molecular basis of the association between transcription regulators nuclear respiratory factor 1 and inhibitor of DNA binding protein 3 and the development of microvascular lesions. *Microvasc Res* 141:104337
11. Das JK, Deoraj A, Roy D, Felty Q (2022) Brain infiltration of breast cancer stem cells is facilitated by paracrine signaling by inhibitor of differentiation 3 to nuclear respiratory factor 1. *J Cancer Res Clin Oncol*. [cited 2023 Feb 18];148(10):2881–91. Available from: <https://link.springer.com/article/10.1007/s00432-022-04026-w>
12. Das JK, Felty Q (2015) Microvascular lesions by estrogen-induced ID3: Its implications in cerebral and cardiorenal vascular disease. *J*

- Mol Neurosci. [cited 2023 Feb 18];55(3):618–31. Available from: <https://link.springer.com/article/10.1007/s12031-014-0401-9>
13. Das JK, Felty Q (2014) PCB153-induced overexpression of ID3 contributes to the development of microvascular lesions. *PLoS One*. [cited 2023 Feb 18];9(8):e104159. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0104159>
 14. Chao AC, Chen CH, Chang SH, Huang CT, Hwang WC, Yang DI (2019) Id1 and sonic hedgehog mediate cell cycle reentry and apoptosis induced by amyloid beta-peptide in post-mitotic cortical neurons. *Mol Neurobiol*. [cited 2023 Feb 18];56(1):465–89. Available from: <https://link.springer.com/article/10.1007/s12035-018-1098-5>
 15. Chen S der, Yang JL, Lin YC, Chao AC, Yang DI (2020) Emerging roles of inhibitor of differentiation-1 in Alzheimer's disease: Cell cycle reentry and beyond. *Cells* 9:1746. [cited 2023 Feb 18];9(7):1746. Available from: <https://www.mdpi.com/2073-4409/9/7/1746/htm>
 16. Preciados M, Yoo C, Roy D (2016) Estrogenic endocrine disrupting chemicals influencing NRF1 regulated gene networks in the development of complex human brain diseases. *Int J Mol Sci* 17:2086. [cited 2023 Feb 18];17(12):2086. Available from: <https://www.mdpi.com/1422-0067/17/12/2086/htm>
 17. Anne UB, Urfer R (2017) Identification of a nuclear respiratory factor 1 recognition motif in the apolipoprotein E variant APOE4 linked to Alzheimer's Disease. *Sci Rep* 7:1. [cited 2023 Feb 18];7(1):1–8. Available from: <https://www.nature.com/articles/srep40668>
 18. Cooper GF, Herskovits E (1991) A Bayesian Method for constructing Bayesian belief networks from databases. In: *Uncertainty proceedings*, pp 86–94
 19. Yoo C, Brilz EM (2009) The five-gene-network data analysis with local causal discovery algorithm using causal Bayesian networks. *Ann N Y Acad Sci*. [cited 2023 Mar 2];1158:93–101. Available from: <https://pubmed.ncbi.nlm.nih.gov/19348635/>
 20. Bhawe K, Das JK, Yoo C, Felty Q, Gong Z, Deoraj A, Liuzzi JP, Ehtesham NZ, et al (2022) Nuclear respiratory factor 1 transcriptomic signatures as prognostic indicators of recurring aggressive mesenchymal glioblastoma and resistance to therapy in White American females. *J Cancer Res Clin Oncol*. [cited 2023 Feb 24];148(7):1641–82. Available from: <https://pubmed.ncbi.nlm.nih.gov/35441887/>
 21. Yoo C, Gonzalez E, Gong Z, Roy D (2022) A better mechanistic understanding of big data through an order search using Causal Bayesian Networks. *Big Data Cogn Comput* 6(2):56. <https://doi.org/10.3390/bdcc6020056>
 22. Kövari E, Herrmann FR, Hof PR, Bouras C (2013) The relationship between cerebral amyloid angiopathy and cortical microinfarcts in brain ageing and Alzheimer's disease. *Neuropathol Appl Neurobiol* [Internet] 39(5):498. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3637988/>
 23. Okamoto K, Amari M, Ikeda M, Fukuda T, Suzuki K, Takatama M (2022) A comparison of cerebral amyloid angiopathy in the cerebellum and CAA-positive occipital lobe of 60 brains from routine autopsies. *Neuropathology*. [cited 2023 Feb 18];42(6):483–7. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1111/neup.12838>
 24. Merlini M, Wanner D, Nitsch RM (2016) Tau pathology-dependent remodelling of cerebral arteries precedes Alzheimer's disease-related microvascular cerebral amyloid angiopathy. *Acta Neuropathol*. [cited 2022 Jun 2];131(5):737–52. Available from: <https://pubmed.ncbi.nlm.nih.gov/26988843/>
 25. Wharton SB, Wang D, Parikh C, Matthews FE, Brayne C, Ince PG (2019) Epidemiological pathology of A β deposition in the ageing brain in CFAS: addition of multiple A β -derived measures does not improve dementia assessment using logistic regression and machine learning approaches. *Acta Neuropathol Commun*. [cited 2023 Feb 18];7(1):198. Available from: <https://actaneurocomms.biomedcentral.com/articles/10.1186/s40478-019-0858-4>
 26. Mao X, Qin X, Li L, Zhou J, Zhou M, Li X, Xu Y, Yuan L, et al (2018) A 15-long non-coding RNA signature to improve prognosis prediction of cervical squamous cell carcinoma. *Gynecol Oncol*. [cited 2023 Feb 18];149(1):181–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/29525275/>
 27. Ma Y, Luo T, Dong D, Wu X, Wang Y (2018) Characterization of long non-coding RNAs to reveal potential prognostic biomarkers in hepatocellular carcinoma. *Gene*. [cited 2023 Feb 18];663:148–56. Available from: <https://pubmed.ncbi.nlm.nih.gov/29684484/>
 28. Braverman NE, D'Agostino MD, MacLean GE (2013) Peroxisome biogenesis disorders: Biological, clinical and pathophysiological perspectives. *Dev Disabil Res Rev*. [cited 2023 Feb 18];17(3):187–96. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/ddrr.1113>
 29. Blankestijn M, Bloks VW, Struik D, Huijckman N, Kloosterhuis N, Wolters JC et al (2022) Mice with a deficiency in Peroxisomal Membrane Protein 4 (PXMP4) display mild changes in hepatic lipid metabolism. *Sci Rep* [Internet] 12(1):2512. <https://doi.org/10.1038/s41598-022-06479-y>
 30. Bhawe K, Das JK, Yoo C, Felty Q, Gong Z, Deoraj A, Liuzzi JP, Ehtesham NZ, et al (2022) Nuclear respiratory factor 1 transcriptomic signatures as prognostic indicators of recurring aggressive mesenchymal glioblastoma and resistance to therapy in White American females. *J Cancer Res Clin Oncol* [cited 2023 Feb 18];148(7):1641–82. Available from: <https://pubmed.ncbi.nlm.nih.gov/35441887/>
 31. Ramos J, Yoo C, Felty Q, Gong Z, Liuzzi JP, Poppiti R, Thakur IS, Goel R, et al (2020) Sensitivity to differential NRF1 gene signatures contributes to breast cancer disparities. *J Cancer Res Clin Oncol*. [cited 2023 Feb 18];146(11):2777–815. Available from: <https://link.springer.com/article/10.1007/s00432-020-03320-9>
 32. Yoo C (2012) Bayesian Method for Causal Discovery of Latent-Variable Models from a Mixture of Experimental and Observational Data. *Comput Stat Data Anal*. [cited 2023 Feb 18];56(7):2183–205. Available from: <https://pubmed.ncbi.nlm.nih.gov/32831439/>
 33. Zhao N, Liu CC, van Ingelgom AJ, Linares C, Kurti A, Knight JA, Heckman MG, Diehl NN, et al (2018) APOE ϵ 2 is associated with increased tau pathology in primary tauopathy. *Nat Commun*. [cited 2023 Feb 18];9(1). Available from: <https://pubmed.ncbi.nlm.nih.gov/30348994/>
 34. Han BH, Zhou ML, Johnson AW, Singh I, Liao F, Vellimana AK, Nelson JW, Milner E, et al (2015) Contribution of reactive oxygen species to cerebral amyloid angiopathy, vasomotor dysfunction, and microhemorrhage in aged Tg2576 mice. *Proc Natl Acad Sci USA*. [cited 2022 Jun 2];112(8):E881–90. Available from: <https://pubmed.ncbi.nlm.nih.gov/25675483/>
 35. Davis J, Cribbs DH, Cotman CW, van Nostrand WE (1999) Pathogenic amyloid beta-protein induces apoptosis in cultured human cerebrovascular smooth muscle cells. *Amyloid*. [cited 2022 Jun 2];6(3):157–64. Available from: <https://pubmed.ncbi.nlm.nih.gov/10524279/>
 36. Fossati S, Cam J, Meyerson J, Mezhericher E, Romero IA, CouraudWeksler POBB, Ghiso J et al (2010) Differential activation of mitochondrial apoptotic pathways by vasculotropic amyloid- β variants in cells composing the cerebral vessel walls. *FASEB J* 24(1):229–241
 37. Jäkel L, de Kort AM, Klijn CJM, Schreuder FHBM, Verbeek MM (2022) Prevalence of cerebral amyloid angiopathy: A systematic review and meta-analysis. *Alzheimers Dement*. [cited 2023 Feb 18];18(1):10–28. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/alz.12366>

38. Nooraei MS, Noori-Zadeh A, Darabi S, Rajaei F, Golmohammadi Z, Abbaszadeh HA (2018) Low level of autophagy-related gene 10 (ATG10) expression in the 6-hydroxydopamine rat model of Parkinson's disease. *Iran Biomed J*. [cited 2023 Feb 18];22(1):15–21. Available from: <https://pubmed.ncbi.nlm.nih.gov/28734275/>
39. Rhee SY (2017) Hypoglycemia and dementia. *Endocrinol Metab (Seoul)*. [cited 2023 Feb 18];32(2):195–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/28685510/>
40. Dutta A, Yang Y, Le BT, Zhang Y, Abdel-Wahab O, Zang C, Mohi G (2021) U2af1 is required for survival and function of hematopoietic stem/progenitor cells. *Leukemia*. [cited 2023 Feb 18];35(8):2382–98. Available from: <https://pubmed.ncbi.nlm.nih.gov/33414485/>
41. Chen C, Zhou P, Zhang Z, Liu Y (2022) U2AF1 mutation connects DNA damage to the alternative splicing of RAD51 in lung adenocarcinomas. *Clin Exp Pharmacol Physiol*. [cited 2023 Feb 18];49(7):740–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/35434831/>
42. Chatterjee S, Burns TF (2017) Targeting heat shock proteins in cancer: a promising therapeutic approach. *Int J Mol Sci* [Internet] 18(9). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5618627/>
43. Yi L, Gu YH, Wang XL, An LZ, Xie XD, Shao W, Ma LY, Fang JR, et al (2016) Association of ACE, ACE2 and UTS2 Polymorphisms with Essential Hypertension in Han and Dongxiang Populations from North-western China. <https://doi.org/10.1177/147323000603400306>. [cited 2023 Feb 18];34(3):272–83. Available from: <https://journals.sagepub.com/doi/10.1177/147323000603400306>
44. Zimdahl H, Haupt A, Brendel M, Bour L, Machicao F, Salsali A, Broedl UC, Woerle H-J et al (2017) Influence of common polymorphisms in the SLC5A2 gene on metabolic traits in subjects at increased risk of diabetes and on response to empagliflozin treatment in patients with diabetes. *Pharmacogenet Genomics* 27(4):135–142
45. Echevarría L, Clemente P, Hernández-Sierra R, Gallardo ME, Fernández-Moreno MA, Garesse R (2014) Glutamyl-tRNA^{Gln} amidotransferase is essential for mammalian mitochondrial translation in vivo. *Biochem J*. [cited 2023 Feb 19];460(1):91–101. Available from: <https://pubmed.ncbi.nlm.nih.gov/24579914/>
46. Coskun PE, Wyrembak J, Derbereva O, Melkonian G, Doran E, Lott IT, Head E, Cotman CW, et al (2010) Systemic mitochondrial dysfunction and the etiology of Alzheimer's disease and down syndrome dementia. *J Alzheimers Dis*. [cited 2023 Feb 19];20 Suppl 2(0 2). Available from: <https://pubmed.ncbi.nlm.nih.gov/20463402/>
47. Guo C, Chen M, Ma W, Cai B, Xu Y, Zhong Y, Zhou C (2020) Growth differentiation factor 9 inhibits vascular endothelial growth factor expression in human granulosa cells. *Gynecol Endocrinol*. [cited 2023 Feb 19];36(10):907–11. Available from: <https://pubmed.ncbi.nlm.nih.gov/31996061/>
48. Wang S, Guo N, Li S, He Y, Zheng D, Li L et al (2021) EZH2 dynamically associates with non-coding RNAs in mouse hearts after acute angiotensin II treatment. *Front Cardiovasc Med* [Internet] 8:585691. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7959742/>
49. Cacabelos R, Torrellas C (2015) Epigenetics of aging and Alzheimer's disease: implications for pharmacogenomics and drug response. *Int J Mol Sci* [Internet] 16(12):30483. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4691177/>
50. Sokol AM, Sztolsztener ME, Wasilewski M, Heinz E, Chacinska A (2014) Mitochondrial protein translocases for survival and wellbeing. *FEBS Lett*. [cited 2023 Feb 19];588(15):2484–95. Available from: <https://pubmed.ncbi.nlm.nih.gov/24866464/>
51. Raghuram V, Weber S, Raber J, Chen DH, Bird TD, Maylie J, Adelman JP (2017) Assessment of mutations in KCNN2 and ZNF135 to patient neurological symptoms. *Neuroreport*. [cited 2023 Feb 19];28(7):375–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/28240725/>
52. Schroer RJ, Holden KR, Tarpey PS, Mathews MG, Griesemer DA, Friez MJ et al (2010) Natural history of Christianson syndrome. *Am J Med Genet A* [Internet] 0(11):2775. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3698558/>
53. Kumar PL, James PF (2015) Identification and characterization of methylation-dependent/independent DNA regulatory elements in the human SLC9B1 gene. *Gene*. [cited 2023 Feb 19];561(2):235–48. Available from: <https://pubmed.ncbi.nlm.nih.gov/25701605/>
54. Liu T, Jiang W, Han D, Yu L (2012) DNAJC25 is downregulated in hepatocellular carcinoma and is a novel tumor suppressor gene. *Oncol Lett*. [cited 2023 Feb 19];4(6):1274–80. Available from: <https://pubmed.ncbi.nlm.nih.gov/23205125/>
55. McCarron MO, Nicoll JAR, Stewart J, Ironside JW, Mann DMA, Love S, Graham DI, Dewar D (1999) The apolipoprotein E epsilon2 allele and the pathological features in cerebral amyloid angiopathy-related hemorrhage. *J Neuropathol Exp Neurol*. [cited 2023 Feb 19];58(7):711–8. Available from: <https://pubmed.ncbi.nlm.nih.gov/10411341/>
56. Kirshner HS, Bradshaw M (2015) The inflammatory form of cerebral amyloid angiopathy or “cerebral amyloid angiopathy-related inflammation” (CAARI). *Curr Neurol Neurosci Rep*. [cited 2023 Feb 19];15(8). Available from: <https://pubmed.ncbi.nlm.nih.gov/26096511/>
57. Sasahira T, Kurihara-Shimomura M, Shimomura H, Bosserhoff AK, Kirita T (2021) Identification of oral squamous cell carcinoma markers MUC2 and SPRR1B downstream of TANGO. *J Cancer Res Clin Oncol*. [cited 2023 Feb 19];147(6):1659–72. Available from: <https://pubmed.ncbi.nlm.nih.gov/33620575/>
58. Ma T, Wu FH, Wu HX, Fa Q, Chen Y (2022) Long non-coding RNA MCM3AP-AS1: A crucial role in human malignancies. *Pathol Oncol Res*. [cited 2023 Feb 19];28. Available from: <https://pubmed.ncbi.nlm.nih.gov/35783356/>
59. Chao AC, Chen CH, Wu MH, Hou BY, Yang DI (2020) Roles of Id1/HIF-1 and CDK5/HIF-1 in cell cycle reentry induced by amyloid-beta peptide in post-mitotic cortical neuron. *Biochim Biophys Acta Mol Cell Res*. [cited 2023 Feb 19];1867(4). Available from: <https://pubmed.ncbi.nlm.nih.gov/31884068/>
60. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. [cited 2023 Feb 24];29(1):15–21. Available from: <https://pubmed.ncbi.nlm.nih.gov/23104886/>
61. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. [cited 2023 Feb 24];30(7):923–30. Available from: <https://pubmed.ncbi.nlm.nih.gov/24227677/>
62. Law CW, Chen Y, Shi W, Smyth GK (2014) Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. [cited 2023 Feb 24];15(2):1–17. Available from: <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-2-r29>
63. Park SB, Hwang KT, Chung CK, Roy D, Yoo C (2020) Causal Bayesian gene networks associated with bone, brain and lung metastasis of breast cancer. *Clin Exp Metastasis*. [cited 2023 Feb 24];37(6):657–74. Available from: <https://pubmed.ncbi.nlm.nih.gov/33083937/>
64. Wang S, Qaisar U, Yin X, Grammas P (2012) Gene expression profiling in Alzheimer's disease brain microvessels. *J Alzheimers Dis*. [cited 2023 Feb 20];31(1):193–205. Available from: <https://pubmed.ncbi.nlm.nih.gov/22531426/>
65. Sartor MA, Leikauf GD, Medvedovic M (2009) LRpath: a logistic regression approach for identifying enriched biological groups in gene expression data. *Bioinformatics*. [cited 2023 Feb 20];25(2):211–7. Available from: <https://pubmed.ncbi.nlm.nih.gov/19038984/>
66. Liu J (2010) Mechanotransduction in endothelial cells: Cell growth, angiogenesis and wound healing [Doctoral dissertation, Ohio State University]. OhioLINK Electronic Theses and Dissertations Center. http://rave.ohiolink.edu/etdc/view?acc_num=osu1274392778

67. Jambusaria A, Hong Z, Zhang L, Srivastava S, Jana A, Toth PT, Dai Y, Malik AB, et al (2020) Endothelial heterogeneity across distinct vascular beds during homeostasis and inflammation. *Elife*. [cited 2022 Jun 2];9. Available from: <https://pubmed.ncbi.nlm.nih.gov/31944177/>
68. Paik DT, Tian L, Williams IM, Rhee S, Zhang H, Liu C, Mishra R, Wu SM, et al (2022) Single-cell RNA-seq unveils unique transcriptomic signatures of organ-specific endothelial cells. *Circulation*. [cited 2022 Jun 2];142(19):1848. Available from: <https://pubmed.ncbi.nlm.nih.gov/37658053/>
69. Paik DT, Tian L, Williams IM, Rhee S, Zhang H, Liu C et al (2020) Single-cell RNA-seq unveils unique transcriptomic signatures of organ-specific endothelial cells. *Circulation* [Internet] 142(19):1848. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7658053/>
70. Nemzek JA, Hodges AP, He Y (2015) Bayesian network analysis of multi-compartmentalized immune responses in a murine model of sepsis and direct lung injury. *BMC Res Notes* 8:516
71. Park SB, Hwang KT, Chung CK, Roy D, Yoo C. Causal Bayesian gene networks associated with bone, brain and lung metastasis of breast cancer. *Clin Exp Metastasis* [Internet]. 2020 Dec 1 [cited 2023 Feb 19];37(6):657–74. Available from: <https://pubmed.ncbi.nlm.nih.gov/33083937/>
72. Kunkle BW, Yoo C, Roy D (2013) Reverse engineering of modified genes by Bayesian network analysis defines molecular determinants critical to the development of glioblastoma. *PLoS One*. [cited 2023 Feb 24];8(5):e64140. Available from: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0064140>
73. Ramos J, Das J, Felty Q, Yoo C, Poppiti R, Murrell D, Foster PJ, Roy D (2018) NRF1 motif sequence-enriched genes involved in ER/PR-ve HER2 +ve breast cancer signaling pathways. *Breast Cancer Res Treat*. [cited 2023 Feb 24];172(2):469–85. Available from: <https://pubmed.ncbi.nlm.nih.gov/30128822/>
74. Lee Y-S, Krishnan A, Oughtred R, Rust J, Chang CS, Ryu J, Kristensen VN, Dolinski K, Theesfeld CL, Troyanskaya OG (2019) A computational framework for Genomewide characterization of the human disease landscape. *Cell Syst* 8(2):152–162. <https://doi.org/10.1016/j.cels.2018.12.010>
75. Jäkel L, De Kort AM, Klijn CJM, Schreuder FHBM, Verbeek MM (2022) Prevalence of cerebral amyloid angiopathy: A systematic review and meta-analysis. *Alzheimers Dement*. [cited 2023 Feb 24];18(1):10–28. Available from: <https://onlinelibrary.wiley.com/doi/full/10.1002/alz.12366>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.