

Article

Not peer-reviewed version

Elucidating the Functional Roles of Long Non-coding RNAs in Alzheimer's Disease

[Zhenyu Huang](#), Qiufen Chen, [Xuechen Mu](#), Zheng An, [Ying Xu](#)*

Posted Date: 10 July 2024

doi: 10.20944/preprints202407.0848.v1

Keywords: Alzheimer's disease; long non-coding RNAs; oxidative stress; Fenton reaction



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Elucidating the Functional Roles of Long Non-coding RNAs in Alzheimer's Disease

Zhenyu Huang ^{1,2}, Qiufen Chen ², Xuechen Mu ^{2,3}, Zheng An ⁴ and Ying Xu ^{1,2,*}

¹ College of Computer Science and Technology, Jilin University, China.

² Systems Biology Lab for Metabolic Reprogramming, Department of Human Genetics and Cell Biology, School of Medicine, Southern University of Science and Technology, Shenzhen, China.

³ School of Mathematics, Jilin University, China.

⁴ Indiana University School of Medicine, USA.

* Correspondence: xuy9@sustech.edu.cn

Abstract: Alzheimer's disease (AD) is a multifaceted neurodegenerative disorder characterized by cognitive decline and neuronal loss, representing a most challenging health issue. We present a computational analysis of transcriptomic data of AD tissues vs. healthy controls, focused on elucidation of functional roles played by long non-coding RNAs (lncRNAs) throughout the AD progression. We first assembled our own lncRNA transcripts from the raw RNA-Seq data generated from 527 samples of dorsolateral prefrontal cortex, resulting in the identification of 31,574 novel lncRNA genes. Based on co-expression analyses between mRNAs and lncRNAs, a co-expression network is constructed. Maximal subnetworks with dense connections are identified as functional clusters. Pathway enrichment analyses are conducted over mRNAs and lncRNAs in each cluster, which serve as the basis for the inference of functional roles played by lncRNAs involved in each of the key steps in an AD development model that we have previously build based on transcriptomic data of protein-encoding genes. Detailed information is presented about the functional roles by lncRNAs in activities related to stress response, reprogrammed metabolism, cell-polarity, and development. Our analyses have also revealed that lncRNAs have discerning power in distinguishing between AD samples of each stage and healthy controls. This study represents the first of its kind.

Keywords: Alzheimer's disease; long non-coding RNAs; oxidative stress; Fenton reaction

1. Introduction

Alzheimer's disease (AD) is a complex disease with numerous pathological changes, such as altered homeostasis of extra- and intracellular pH [1,2], disrupted equilibrium of crucial intracellular ions such as Na^+ and K^+ [3,4], chronic inflammation, heightened oxidative stress [5,6], elevated neurotoxicity, and extensive neuronal apoptosis [7,8]. While considerable attention has been devoted to unraveling the causes and consequences of amyloid-beta ($\text{A}\beta$) plaque development and Tau-based neurofibrillary tangle (NFT) formation in the past three decades, recent studies have raised doubts regarding their proposed central roles in the reduced cognitive capacities in AD patients [9]. Instead, a preponderance of evidence suggests extensive neuronal death in specific cerebral regions as the key determinant of the cognitive decline in the affected individuals [10]. Although great amounts of omics data and information amassed in the field of AD research, disentangling the intricate causal relationships among these pathological conditions and their relevance to the extensive neuronal loss remains a great challenge. Numerous hypotheses have been put forward regarding the drivers and key mechanisms of the development of AD, including oxidative stress in mitochondria [11], and dysregulated intracellular pH, which can impair the functions of acidic organelles like endosomes and lysosomes, thereby detrimental to the host neurons [12,13]. Additionally, the formation of Tau aggregates can precipitate cell death [14], while overexpression of the amyloid precursor protein (APP) may contribute to $\text{A}\beta$ -associated neuronal death in advanced AD tissues [14].

Like in studies of other human diseases, vast majority of the published work on AD has been protein-centric [15] while the roles played by long non-coding RNAs (lncRNAs) have been largely un- or underexplored [13]. Limited knowledge about lncRNAs in the pathogenesis of AD includes those involved in amyloid formation [16], Tau protein hyperphosphorylation [17], and oxidative-stress response [18]. As a comparison, the functional roles played by lncRNAs under physiological conditions have been extensively studied in brain development, homeostasis, oxidative stress response, plasticity, and evolution [11–13]. It has been well established that 40% of the human lncRNA genes are expressed in the brain, hence there should be no surprise if lncRNAs play important roles in AD pathogenesis.

In this study, we have performed transcript-level assembly of all RNAs from raw RNA-seq data collected from the prefrontal cortex tissues of both healthy individuals and AD patients, which are publicly available [19]. The reason for doing our own assembly of transcripts is that the available transcripts of RNA-genes in AD tissues in the public domain are far from being adequate for meaningful analyses. Specifically, transcripts for only 11,300 lncRNA genes have been detected in AD tissues and made publicly available, which is clearly too low, knowing that overall, 228,048 transcripts for 48,479 lncRNA genes have been detected in healthy human tissues and publicly available. Our assembly resulted in 431,781 transcripts for 55,098 lncRNA genes, having considerably expanded the numbers of both the transcripts and the genes.

It has been well established that AD tissue cells have increased intracellular pH with a pH at roughly 7.037 and decreased extracellular pH at 6.85, compared to the normal ones at ~7.028 [20] and 6.9 [21,22]. We have conducted a computational analysis of the transcriptomic data of AD tissues vs. controls, coupled with computational chemistry analyses, to address two main questions: (1) what are the causes and consequences of AD cells intracellular alkalization? and (2) what are the causes and consequences of AD cells extracellular acidification? Our answers to these questions naturally give rise to a model for AD formation and development (manuscript under review), consisting of the following key steps and illustrated in Figure 1:

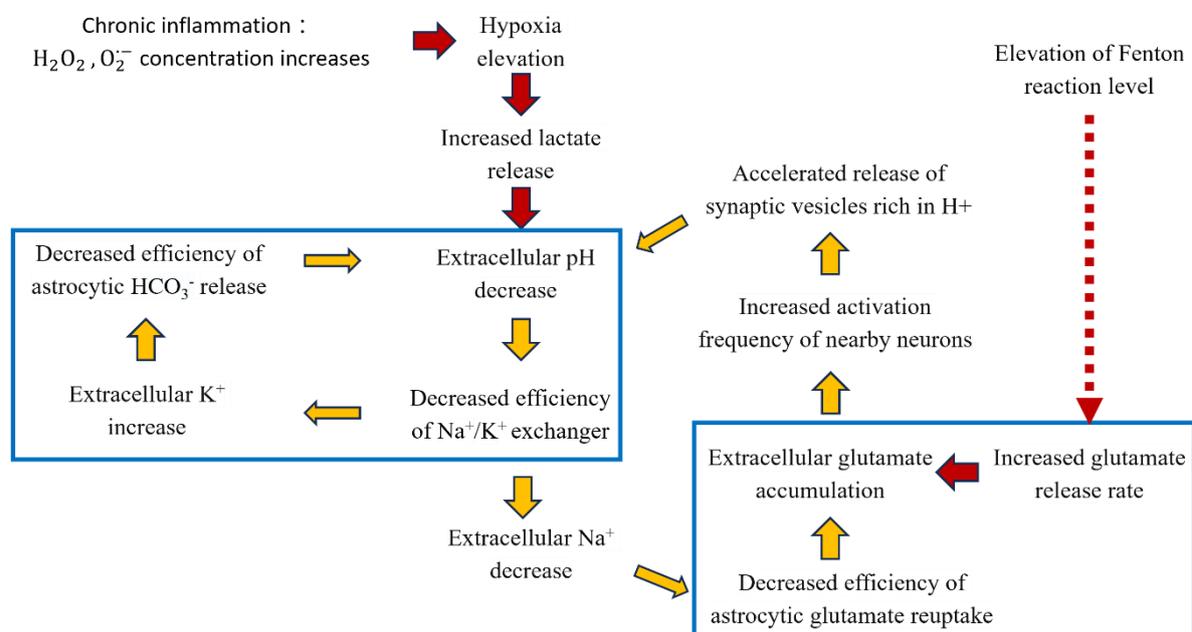
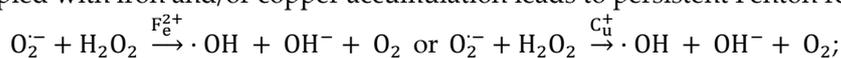
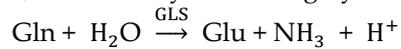


Figure 1. A model for two vicious cycles driven by Fenton reactions throughout the progression of AD.

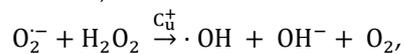
1) chronic inflammation (chemically, the levels of H_2O_2 and O_2^- are significantly increased) coupled with iron and/or copper accumulation leads to persistent Fenton reaction in mitochondria:



2) mitochondrial persistent Fenton reactions drive the pH up, leading to cell death if not neutralized; 3) multiple acidifying metabolic reprograms are induced to produce H^+ to keep the pH stable, with the key ones being hydrolysis of glutamine to glutamate and NH_4^+ , catalyzed by GLS:



and hyperphosphorylation of the Tau proteins bound with microtubules, which produces one H^+ per phosphorylation, resulting in acidic Tau fiber structure [23]; 4) persistently produced Glu are released into the extracellular space; 5) the over-produced and released glutamates drive the neighboring neurons hyperexcited, which gives rise to over-production and release of acidic synaptic vesicles, hence resulting in acidification of the extracellular space; 6) under physiological conditions, the H^+ released by synaptic vesicles will be neutralized by bicarbonates released by nearby astrocytes while under the condition of increasing extracellular acidification, the release rate of bicarbonate by astrocytes could not keep up with the release rate of synaptic vesicles, forming one vicious cycle involving increased extracellular K^+ level and further decreased extracellular pH; 7) as the disease evolves, the extracellular level of glutamate accumulation increases, as a result of a second vicious cycle involving decreased extracellular Na^+ level, further decreased extracellular pH, and reduced ability to clear extracellular glutamates; 8) a key response to progressively increasing extracellular acidification, cells increase their release rate of C_u^+ , resulting in extracellular Fenton reaction:



which helps to slow down the extracellular acidification in two ways: the persistent production of OH^- and the formation of alkaline $A\beta$ plaques [24] as the result of interactions between $\cdot\text{OH}$ and $A\beta$ monomers [25].

Our analyses have revealed that extracellular acidosis represents the leading cause for neuronal cell death in AD, followed by the formation of $A\beta$ plaques only in advanced stage AD while it plays a role in slowing down the extracellular acidosis in the early phase of the disease (manuscript under review).

Compared to this summary of our previous work, the current study focuses on the functional roles played by lncRNAs throughout the disease development of AD. To study the progression of AD, we have developed a pseudo-time course based on the AD tissue samples under study, arranged in the increasing order of the sum of the absolute values of the differential expressions of genes in a disease tissue vs. controls. Using this ordered list, we have made the following observations regarding lncRNA gene's involvement in AD pathogenesis:

1. Expressed lncRNAs are primarily involved in the upregulated functions in AD tissues, ranging from immune activities, metabolisms to cell polarity and stress-responses, which seem to play driving roles of the disease progression, while only a few lncRNAs are involved in downregulated functions;

2. The most enriched pathways by lncRNAs expressed in AD tissues are relevant to oxidative stress;

3. lncRNA-mediated reprogrammed metabolisms are involved in alleviating intracellular alkaline stress and extracellular acidic stress; and

4. lncRNA as biomarkers have strong predictive power in distinguishing (early-stage) AD patients from healthy controls.

To the best of our knowledge, this is the first large-scale analysis of the functional roles played by lncRNAs throughout AD development.

2. Results

2.1. Identification and Characterization of Novel lncRNAs

We have assembled the transcriptome from raw RNA-seq data of 527 AD and control tissue samples, and then removed transcripts for protein genes (Data and Methods). This resulted in 55,098 predicted lncRNAs. The clinical information of these samples was summarized in [Supplementary Table S1](#).

Of these lncRNAs, 31,574 are of high-quality and novel, namely not present in the GENCODE database [26], as summarized in Figure 2A and detailed in Data S1. 41.79% of these transcripts were in intergenic regions, 46.24% in the intronic regions, and 11.97% are antisense of protein-encoding genes (Figure 2B), suggesting their possible roles in transcription regulation of the relevant protein genes.

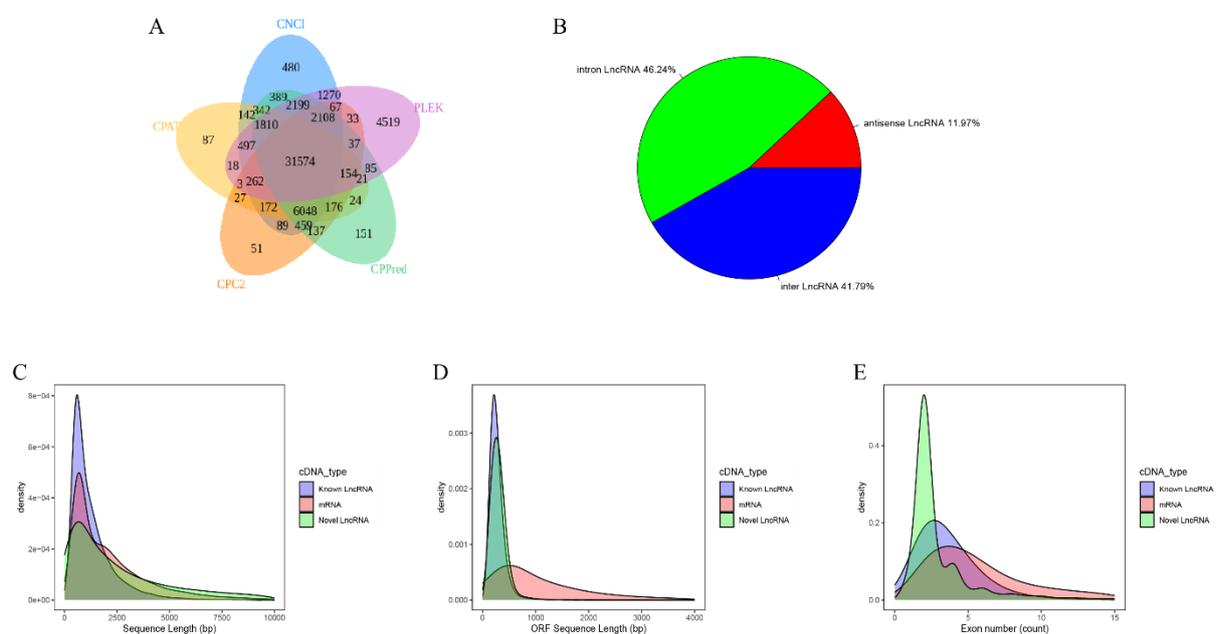


Figure 2. Prediction of novel lncRNAs in AD tissues. **A.** Venn analysis of the predicted novel lncRNAs using five software: CNCI, CPC, CPP2, PLEK and CAPT. **B.** Classification of predicted novel lncRNAs. **C.** Exon numbers in protein-coding genes, known lncRNA genes, and novel lncRNAs. **D.** ORF length distributions of protein coding genes, known lncRNAs, and novel lncRNAs. **E.** Sequence lengths of protein-coding genes, known lncRNAs, and novel lncRNAs.

Previous studies have shown that lncRNAs in mammals are shorter in length and have fewer exons than protein-coding genes [27–29]. To examine whether the lncRNAs expressed in AD samples have the same characteristics, we have compared the structures of our assembled lncRNA transcripts (Figure 2C-E), using the two-sided Wilcoxon test [30]. We note that the lncRNAs tend to have considerably shorter average lengths than mRNAs, 1,521 bp vs. 1,842 bp with p -value $< 2.22e-16$ as shown in Figures 2C and S1. And they tend to reside in shorter open reading frames (ORFs) vs. protein genes, having averages 279 bp vs 759 bp with p -value $< 2.22e-16$ as shown in Figures 2D and S1; and each encoded by smaller numbers of exons, two vs. five with p -value $< 2.22e-16$ (Figures 2E and S1).

2.2. lncRNAs' Roles in AD Formation and Development

To study how the phenotypic changes of the disease with the progression of AD, consider bin1, bin2, bin3, bin4 (Figures 3A). Construct a lncRNA-mRNA interaction network over samples in each bin (see METHODS), resulting in 45, 54, 47, and 35 clusters in the four bins, respectively. This is followed by a pathway enrichment analysis over samples in each bin, resulting in a total of 8,000 distinct pathways.

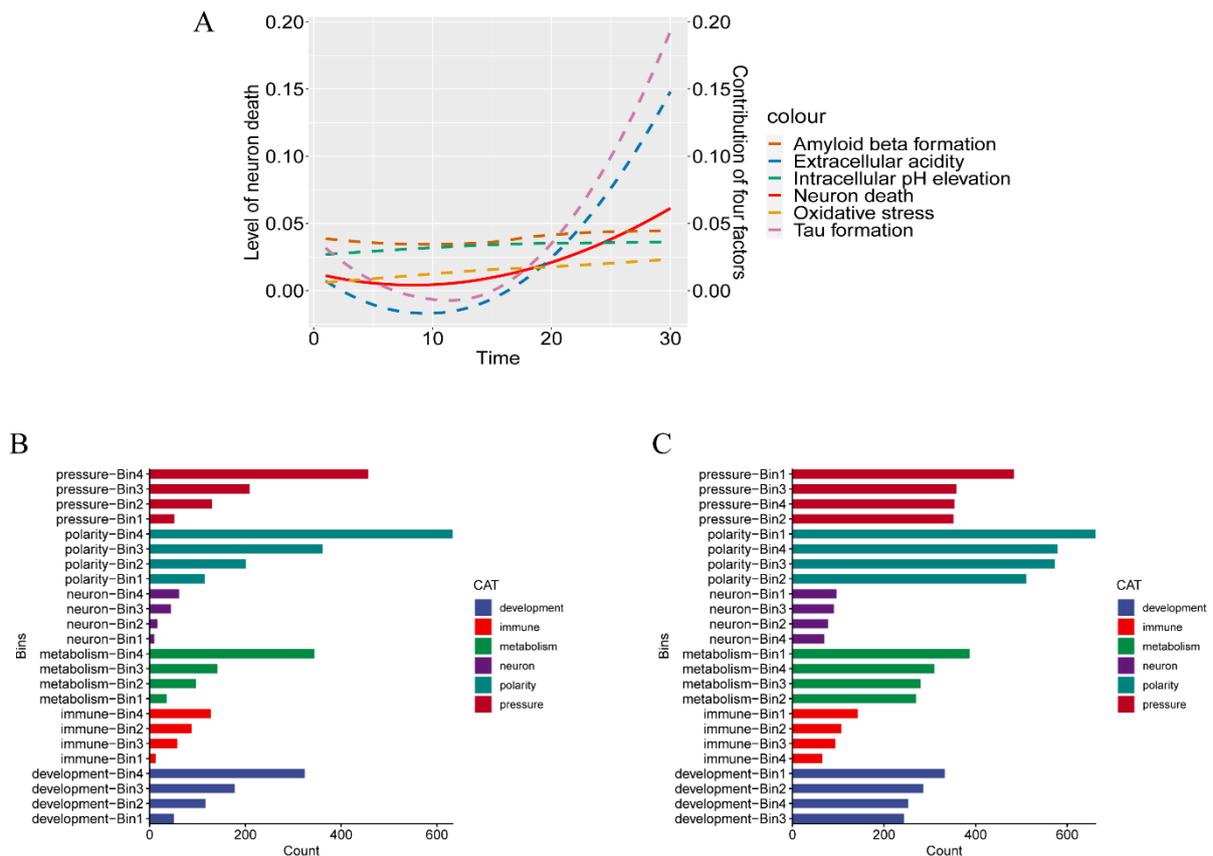


Figure 3. Statistics related to the pathological progression of lncRNAs in AD. **A.** The relative changes in the levels of A β plaques, Tau fibers, intracellular alkalinity, extracellular acidity, oxidative stress, and apoptosis compared to controls throughout the disease progression. **B.** Category statistics of different upregulated pathways in different bins. **C.** Category statistics of different downregulated pathways in different bins.

We grouped these pathways into six categories: development-related, immune activity, metabolism, neural function, cell polarity, and stress response. We note: (1) the number of downregulated pathways in each category remains generally stable across the four bins, namely throughout the progression of an AD; (2) the numbers of upregulated pathways in the categories of development-related, immune activity, metabolism, cell polarity, and stress response each increase with the disease progression, while those in neural function remain relatively stable (Figures 3B-C). Most of the developmental pathways consist of muscle, cell morphology, and a few others. Consequently, we hypothesize that lncRNAs are primarily involved in cell polarity, stress response, immunity, and metabolism in AD. In the following, we investigate how lncRNAs may involve these pathways throughout the AD progression.

Figure 4 summarizes the key components of our protein-centric model for AD development (manuscript under review). Here we study how lncRNAs contribute to the key components in this model along the disease progression from bin1 through bin4. We have quantified the contribution by each pathway to AD progression from bin1 to bin4, where the level of contribution is defined using the discerning score by each feature (see METHODS).

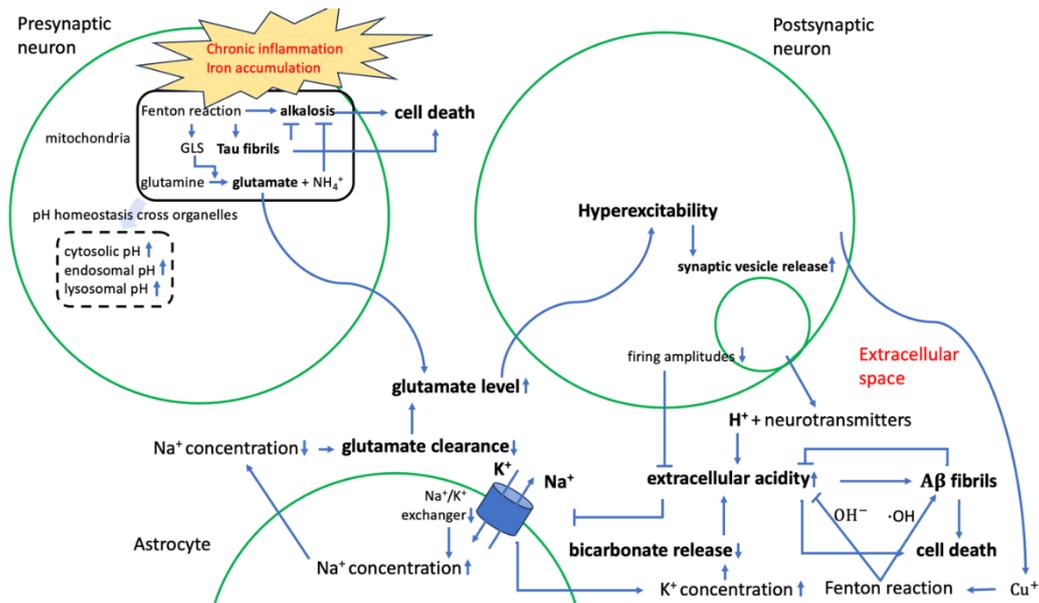


Figure 4. A schematic of our model, composed of key events and logical relationships, where each arrow represents a causal relationship.

2.2.1. Contributions to key AD Phenotypes Via Cell-Polarity Changes

We note that cell-polarity related pathways are consistently making the highest levels of contribution to AD progression across all four bins (Figure 5A and S2, Table S2). The most contributing cell-polarity pathways are filament-bundle assembly, microtubule polymerization or depolymerization, metal ion transport, and transport of acidic organic compounds (Table S2).

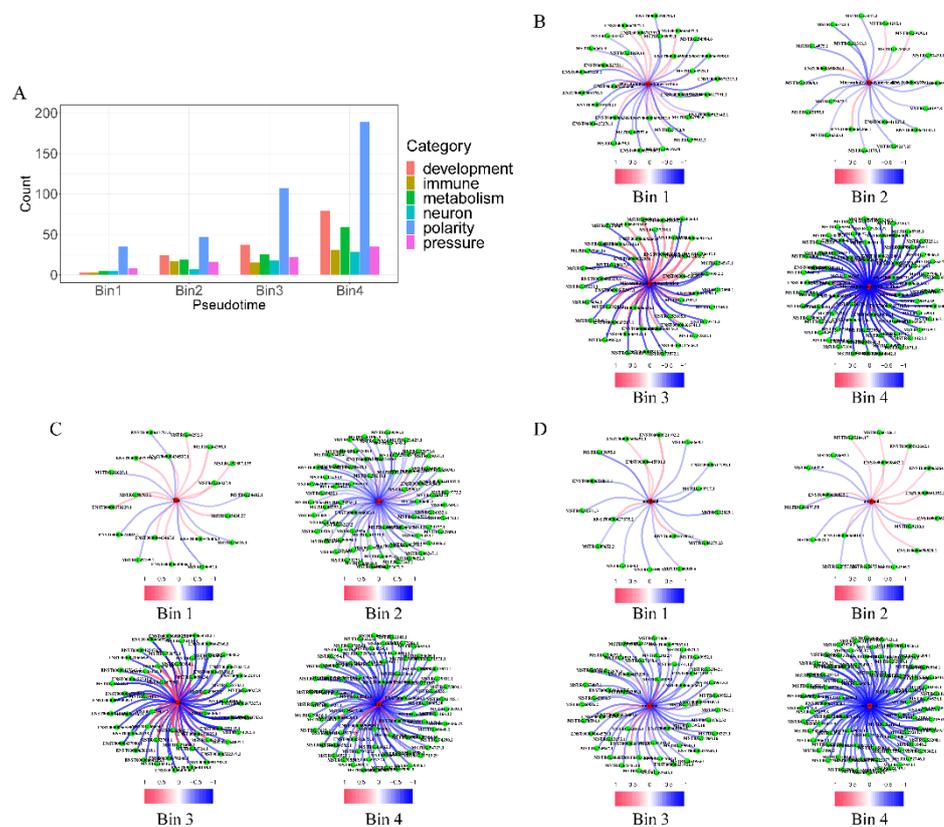


Figure 5. LncRNAs contributing to key AD phenotypes through cell-polarity changes. **A.** Statistics of the top 20% of pathways for each group of pathways in each bin. **B.** The regulatory trends of lncRNAs

associated with microtubule depolymerization from bin1 through bin4. C. The numbers of lncRNAs involved in Tau fiber formation across the four bins. D. The numbers of lncRNAs involved in A β plaque formation across the four bins.

In AD, filament bundle assembly refers to the formation of intracellular Tau fibril bundles and the formation of extracellular A β plaques [14]. The process of intracellular Tau fiber formation starts from the disassociation between the Tau proteins and microtubules that they bind, resulting in free Tau monomers in the intracellular space and then the formation of Tau aggregates.

We note that the number of lncRNAs involved in microtubule depolymerization via freeing up Tau proteins ranges 14, 5, 208, to 185 from bin1 through bin4, respectively (Figure 5B, Table S3). Examples include lncRNA ENST00000666526.1, which enhances the transcription of KIFC3, involved in microtubule transport and positioning [31], and STMN1 that promotes microtubule depolymerization [32]. Interestingly, lncRNAs are also involved in inhibiting microtubule depolymerization, with the number of such lncRNAs ranging from 18, 30, 278, to 3,387 across the four bins, respectively (Figure 5B, Table S3). One example is lncRNA MSTRG.54738.1 involved in suppressing the expressions of STMN1 and SPAST, where the latter is a microtubule-severing protein involved in regulating microtubule dynamics (Table S3) [33]. We postulate that the differences in the numbers of lncRNAs involved in promoting vs. those in inhibiting microtubule depolymerization reflect a shift in the homeostasis of cellular populations of microtubules.

We also note that the number of lncRNAs involved in Tau-fiber formation ranges from 17, 11, 170, to 182 from bin1 to bin4 while the number of those in inhibiting Tau-fiber formation ranges from 21, 137, 175, to 3,056, respectively (Figures 5C, Table S4). For example, lncRNA MSTRG.44295.1 exhibits the strongest promotion of BRSK2 transcripts across the four bins (Table S4). Interestingly, BRSK2 is also heavily suppressed by lncRNAs, with MSTRG.61027.1 showing the greatest inhibitory potency (Table S4). Again, these data suggest that there are opposing forces in promoting and inhibiting the formation of Tau fibers, which is consistent with published studies [34] and our own data (Figure 5C, S3). To further understand this, a pathway-enrichment analysis over mRNAs correlated with lncRNAs involved in Tau fiber formation, as well as with lncRNAs in inhibiting its formation. We note that (1) pathways correlated with Tau-fiber formation are predominantly involved in intracellular alkaline-stress response; and (2) pathways relevant to inhibiting Tau fibers are neuronal death (Table S4).

Similar observations are made about lncRNAs involvement in the development of A β plaques. Specifically, in the early-stage samples (bin1 and bin2), lncRNAs are mostly involved in promotion of A β plaque formation, while in the advanced stage samples (bin3, bin4), lncRNAs are heavily involved in inhibition of A β plaque formation (Figure 5D). Examples include: (1) lncRNAs in Table S5 are involved in expressing CLU, a protein crucial for A β plaque formation; and (2) lncRNA MSTRG.19468.1 is involved in inhibiting A β plaque formation (Table S5). Our explanation is that the formation of A β plaques serves roles in slowing down the extracellular acidosis process, as the copper-mediated extracellular Fenton reactions produce OH \cdot and the alkaline A β plaque structures as outlined in the Introduction while with the increase in both the density and sizes of A β plaques, they become increasingly toxic to the nearby neurons [24].

2.2.2. Contribution to Stress Generation

The most significant stressors throughout AD development are intracellular alkalization and extracellular acidification, both being the results of mitochondrial Fenton reactions (Figure 4 and manuscript under review). Our previous study [35] has revealed that AD involves at least two classes of Fenton reactions, one in mitochondria, either iron or copper mediated, which we consider as an initial driver of the formation of AD, and one in extracellular space, predominantly copper-mediated, which serves a positive role in slowing-down extracellular acidosis through the formation of A β plaques that become deadly as they continue to accumulate [36].

The basis for Fenton reactions is the over-production and accumulation of H₂O₂ and \cdot O₂⁻ by the local astrocytes. lncRNAs are involved in promotion of H₂O₂ and \cdot O₂⁻ production in bin1 and

bin2 and in inhibition of their production in bin3 and bin4 (Figures 6A and S4). The key proteins involved in H_2O_2 production are COX2, SOD2 and MFN2. LncRNA ENST00000560481.2 is a key promoter of the expressions of these proteins, while lncRNA ENST00000667795.1 is a key inhibitor (Table S6). Similarly, the key proteins involved in $\cdot O_2^-$ production are APP, GNAI2 and MAPT, with lncRNA ENST00000666526.1 being a key promoter (Table S6), while HVCN1, AGT, MAPT, CYBA and APP are involved in inhibiting the production of $\cdot O_2^-$, with lncRNA MSTRG.16333.1 being a key inhibitor (Table S6).

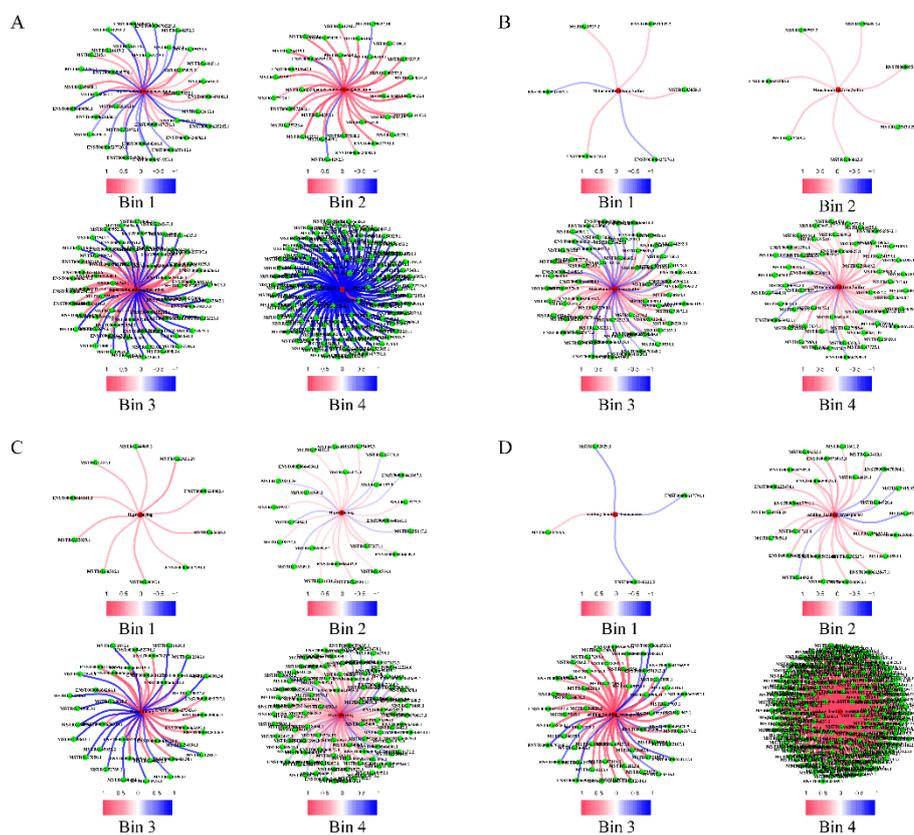


Figure 6. LncRNAs involved in induction of metabolic reprogramming for relieving stresses. **A.** The numbers of lncRNAs involved in superoxide anion generation across the four bins. **B.** The numbers of lncRNAs involved in mitochondrial iron sulfur clustering synthesis across the four bins. **C.** The numbers of lncRNAs involved in H^+ -producing enzymes across the four bins. **D.** The numbers of lncRNAs involved in acid loading transporter across the four bins.

Regarding the roles played by lncRNAs in mitochondrial iron and copper transport, we note: the numbers of lncRNAs involved in iron accumulation in mitochondria are 4, 7, 48, and 54, respectively, while those involved in inhibitory roles are 2, 0, 16, and 5, respectively (Figure 6B, Table S7). Key relevant lncRNAs include MSTRG.45979.1 promoting iron/copper accumulation and MSTRG.21672.6 for inhibiting such accumulation (Figure 6B, Table S7).

In addition, lncRNAs are also involved in promoting extracellular copper release, with their numbers ranging in 113, 29, 260, and 173 from bin1 throughout bin4, and the numbers of lncRNAs involved in copper release are 63, 354, 360, and 3,357 from bin1 to bin4, respectively (Figure S4, Table S8). Among these, MSTRG.24138.1 and MSTRG.16333.1 are examples, one for promoting and the other for inhibiting copper release (Figure S4, Table S8).

Synapse assembly, driven by neuronal firing triggered by extracellular glutamates, leads to the release of H^+ -rich synaptic vesicles, a key contributor to extracellular acidosis (see Introduction) (Figure S4, Table S9). LncRNAs such as MSTRG.43821.1 are predominantly involved in inhibiting synaptic assembly (Table S9).

2.2.3. Contribution to Stress-Response Via Metabolic Reprogramming

Numerous reprogrammed metabolisms are induced in AD tissue cells, responding to intracellular alkalization or extracellular acidification. One class of reprogrammed metabolisms, in which lncRNAs are heavily involved are lipid metabolism, particularly cholesterol metabolism (see Table S2). Specifically, the numbers of lncRNAs involved in promoting cholesterol synthesis are 10, 10, 210, and 246 from bin1 through bin4, respectively, while the numbers of lncRNAs involved in inhibitory roles of cholesterol biosynthesis and metabolism are 14, 60, 245, and 3,644 from bin1 through bin4, respectively (Figure S4, Table S10). This is not surprising knowing that cholesterol has both pro- and anti-inflammatory functions [37].

Knowing that AD cells are stressed with persistent intracellular alkalinity, we have analyzed all the 1,258 enzymes, each catalyzing H⁺-producing reactions [38]. A principal component analysis is conducted over these enzymes across all the AD samples under study. Remarkably, the first principal component, or PC1, across samples in bin4 can explain 0.978 of the variance ratios of the gene expressions of these enzymes (Figure S5), strongly suggesting their consistent roles in alleviating intracellular alkaline stress. LncRNAs are involved in promoting H⁺-generating enzymes from bin1 through bin4, account for 32, 15, 188, 95 in numbers, respectively, while the numbers of lncRNAs involved in inhibiting these enzymes are 5, 17, 205, to 33 from bin1 through bin4 (Figure 6C, Table S11). Among the top 5% of H⁺-producing enzymes contributing to PC1 in each bin, kinases dominate (Table S12). GLS is among these enzymes consistently involved in mitigating the alkalizing stress due to mitochondrial Fenton reactions throughout all bins. Our study found that 50% of the lncRNAs that promote the production of H⁺ enzymes are involved in promoting GLS-related transcripts, such as ENST00000479552.1 (Table S12).

In addition, the number of lncRNAs involved in promoting transporters that acidify the intracellular space (Figure 6D) ranges from 3, 28, 182 to 3,308 from bin1 to bin4, while the number of lncRNAs involved in inhibiting such transporters ranges from 8, 2, 91 to 136 (Table S13), suggesting that how the balance is shifted towards producing more acidic molecules.

Table 1 lists the numbers of lncRNAs with high discerning scores involved in immune activity, metabolic reprogramming, cell polarity, and stress response pathways from bin1 through bin4.

Table 1. The number of lncRNAs with high discerning scores involved in immune activity, metabolic reprogramming, cell polarity, and stress-response from bin1 through bin4.

Pesudotime	Immune activity	Metabolic reprogramming	Cell polarity	Stress response
Bin1	160	339	876	827
Bin2	1069	729	1569	1291
Bin3	863	1362	1522	1421
Bin4	1916	5156	5230	5204

2.3. LncRNA's Discerning Power in Distinguishing AD Samples from Controls

The above analyses have revealed that lncRNAs are involved in the regulation of at least 22 pathways strongly associated with AD phenotypes, as shown in Figure 7A and Table 2. We aim to identify a subset of these lncRNAs that can well distinguish between normal and AD tissues (or tissues of a specific stage of AD), which could potentially be used as blood biomarkers for AD detection. It is noteworthy that compared to proteins, lncRNAs tend to have longer half-lives in blood circulation as they tend to be more resistant to degradation [39–41].

Table 2. AUC scores achieved by lncRNAs in distinguishing between all AD samples vs. controls, late-stage AD samples vs. controls, early-stage AD patients vs. controls, and early-stage vs. late-stage AD samples.

Phenotype	Normal_vs_allAD	Normal_vs_eAD	Normal_vs_aAD	eAD_vs_aAD
Aciding extrusion transporter	0.7124	0.6018	0.8998	0.9499
Aciding loading transporter	0.6448	0.6539	0.825	0.9602

Active oxygen	0.7297	0.7365	0.8673	0.8543
Astrocytes	0.6653	0.7124	0.8029	0.9718
Hydrogen peroxide	0.6981	0.851	0.8412	0.9339
Microglia	0.6669	0.581	0.8614	0.9429
Superoxide anion generation	0.7015	0.823	0.944	0.9416
Bicarbonate transporter	0.6819	0.6675	0.8595	0.9685
Extracellular acidosis	0.6373	0.6425	0.8783	0.878
Na ⁺ /K ⁺ -ATPase	0.7866	0.8256	0.8881	0.9615
Cholesterol	0.6694	0.6871	0.8939	0.9557
Extracellular copper	0.701	0.8783	0.8718	0.9018
Mitochondrial iron sulfur	0.6378	0.5979	0.8022	0.8228
H-producing enzyme	0.6971	0.6662	0.8399	0.8652
Microtubule depolymerization	0.7562	0.8256	0.8198	0.9268
Synapse assembly	0.6909	0.7235	0.8979	0.8967
Amyloid	0.6914	0.6012	0.8855	0.9634
Tau	0.6858	0.6291	0.8549	0.957

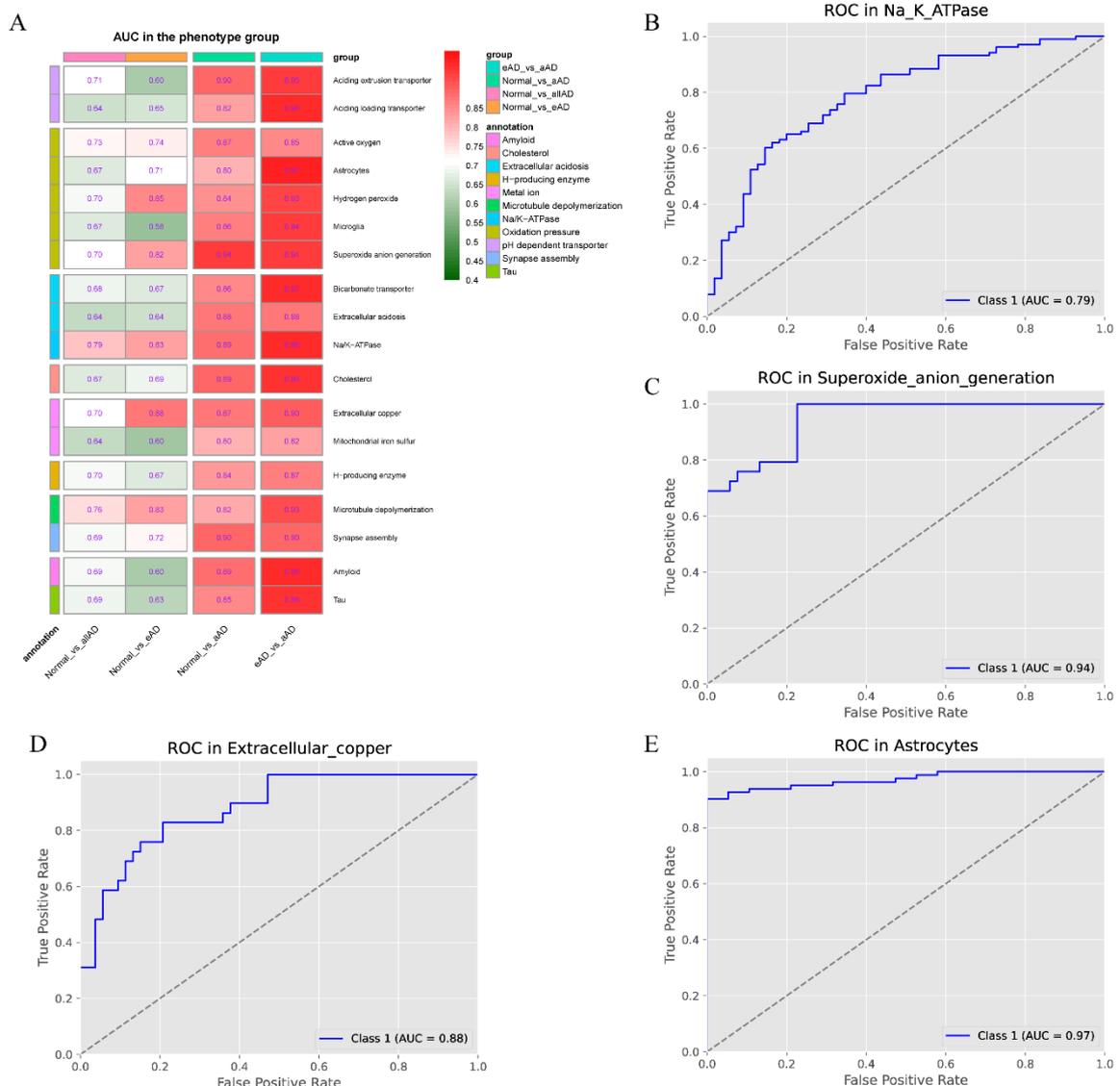


Figure 7. Assessment of lncRNAs' discerning power between AD and control samples. **A.** AUC scores by lncRNAs in distinguishing all AD samples vs. controls, late-stage AD samples vs. controls, earliest-

stage AD samples vs. controls, and early-stage AD samples vs. mid-late-stage AD samples. **B.** ROC curves by lncRNAs involved in Na^+/K^+ -ATPase in distinguishing all AD samples vs. controls. **C.** ROC curves by lncRNAs involved in regulating superoxide anion generation in differentiating late-stage AD samples vs. controls. **D.** ROC curves by lncRNAs regulating extracellular copper in discriminating earliest-stage AD samples vs. controls. **E.** ROC curves by lncRNAs involved in regulating astrocytes in distinguishing early-stage AD samples vs. mid-late-stage AD samples.

Among the lncRNA-containing pathways, lncRNAs involved in regulating Na^+/K^+ -ATPase, suppressed by extracellular acidosis [42], exhibit the best performance in distinguishing normal from AD samples, with $\text{AUC} = 0.7866$ (Figure 7B, Table 2), which is significantly better than $\text{AUC} = 0.655$ achieved by CSF $\text{A}\beta_{40/42}$ prediction [43]. In distinguishing normal from late-stage AD samples, the lncRNAs involved in regulating superoxide anion generation performs the best, with $\text{AUC} = 0.944$ (Figure 7C, Table 2), considerably better than $\text{AUC} = 0.853$ achieved by plasma p-tau181 prediction [44].

Furthermore, in distinguishing normal from the earliest stage AD samples, lncRNAs involved in regulating extracellular copper achieved a performance level at $\text{AUC} = 0.8783$ (Figure 7D, Table 2). In distinguishing the earliest-stage from the late-stage AD samples, the lncRNAs involved in regulating astrocytes achieved $\text{AUC} = 0.9718$ (Figure 7E, Table 2). Detailed information on these lncRNAs can be found in Table S14.

3. Concluding Remarks

We have conducted a computational analysis of the functional roles played by lncRNAs throughout the development of an AD. As the first step of this analysis, we have conducted transcript-level assembly from the raw RNA-seq data, resulting in a total of 55,098 lncRNAs with 30,102 being novel ones.

The basis of our functional analyses of lncRNAs is the pathways enriched by mRNAs having known functional annotations, strongly co-expressed with the lncRNAs, coupled with the predicted functions as *cis* or *trans* regulators.

To elucidate how the functional roles played by lncRNAs change with the progression of the disease, we have defined a distance between the AD samples and the controls, which gives rise to the ordered list of AD samples from the early to the advanced stage. This provides an approximation to the disease progression. Our overall discovery can be summarized as follows:

1. Across all four bins representing different stages of AD progression, cell-polarity related pathways consistently contribute significantly. Notably, filament-bundle assembly, microtubule polymerization or depolymerization, metal ion transport, and transport of acidic organic compounds are among the most contributing pathways, for each of which lncRNAs are actively involved throughout AD progression.

2. Intracellular alkalization, extracellular acidification, and oxidative stress are major stressors and significantly contribute to AD progression. lncRNAs are involved in both the stress generation and stress responses, including iron and copper accumulation as well as the formation of Tau fibers and $\text{A}\beta$ plaques.

3. Intracellular alkalization and extracellular acidification induce multiple acidifying metabolic reprogramming in AD cells for survival. lncRNAs are involved in activating majority of these reprogrammed metabolisms.

4. Figure 4 outlines the steps in our AD model, and lncRNAs are involved in most of these steps.

5. lncRNAs have strong discerning power in distinguishing AD samples from the controls, as well as in distinguishing AD samples of specific stage from the controls.

In summary, this study significantly expands our understanding of how lncRNAs participate in the progression of AD for the first time. The identification and characterization of novel lncRNAs provide valuable insights into their functional roles. Pseudo-time provided by AD samples arranged in a specific order offers highly useful information regarding how various phenotypes of AD change as the disease progresses. These findings may have significant implications for the development of

diagnostic biomarkers and therapeutic targets for AD, paving the way for new avenues in research on lncRNA-based interventions for neurodegenerative diseases.

4. Materials and Methods

4.1. Data

Gene expression data: The raw reads of 529 samples were downloaded from ROSMAP cohort [45], consisting of transcriptomes of 146 MCI (mild cognitive impairment), 193 AD, and 189 region-matched control tissues. All libraries were prepared with the strand-specific RNA-seq paired-end protocol and each on average consisting of 4.28484 million reads. The detailed information about these samples is given in Supplementary Table S1.

Enzymes and reactions: The balanced equation of the chemical reaction catalyzed by each enzyme encoded in the human genome, along with the number of H⁺ produced or consumed, is retrieved from the HumanCyc database [46].

4.2. Methods

4.2.1. RNA-seq Processing and lncRNA Identification

High-quality clean reads were obtained by using the clipping adapters and low-quality reads identified by using SOAPnuke [47]. Each library was individually mapped to the GENCODE genome (version 34) using HISAT2 (version 2.7) [48] with the default parameters. The resulting alignment files were used as input for transcript assembly by using StringTie (version 2.1.1) [27] and the *Homo sapiens* reference annotation of GENCODE (release 34). All StringTie output files were merged into one unified transcriptome using the merge option of StringTie. Gffcompare [27] was employed to compare the resulting transcriptome (GTF format) with the reference annotation. For consistency in nomenclature, we adopt StringTie's default "MSTRG" as the starting point when naming all assembled transcripts.

4.2.2. The Prediction Process Consists of Two Steps

Step 1: screening for protein-coding potentials. Transcript lengths and the number of exons were collected in the basic screening step. Here, transcripts having lengths > 200 nt and at least two exons were selected as lncRNA candidates [29]. The protein-coding potential of a transcript was predicted jointly using five methods: CPC (Coding Potential Calculator) [49], CNCI (Coding-Non-Coding Index) [50], CPAT (Coding Potential Assessment Tool) [51], PLEK [52], and CPPred [53].

Step 2: a candidate is predicted to be a lncRNA if all five methods predict it to be a lncRNA. If the transcript is newly expressed (not in GENCODE), it is assigned one of the following codes: *i* (transcript contained entirely in an intron), *y* (containing a reference within its introns), *p* (not inside but within 2 Kb of a known transcript), or *u* (in an intergenic region). All lncRNAs are then classified into three types: intergenic lncRNAs, antisense lncRNAs, and intronic lncRNAs.

4.2.3. Differential Analysis of All Assembled Transcripts

Differential expression analyses between the (MCI U AD) samples and controls were conducted using the "DESeq2" function [54]. For read-count data, transcripts with $|FC| \geq 1.3$ having a statistical significance p -value < 0.05 were considered differentially expressed transcripts (DETs), where FC is for fold change.

The TPM values of all genes from Ballgown [55] were used to assess correlations between a lncRNA and mRNAs using the Spearman correlation test, where correlation coefficients > 0 with p -values < 0.05 are used as the cutoffs for co-expressed transcript pairs.

4.2.4. Ordering Disease Samples by the Level of Deviation from Control Tissues

A differential expression analysis is conducted on each sample in (MCI \cup AD) compared to the controls, yielding a differential intensity (V) value for each sample, defined as the sum of the absolute values of differential expressions over all DETs with p -value < 0.05 . We aim to arrange the (MCI \cup AD) samples in the increasing order of the V values as an approximation to the disease progression. However, the order so defined is not very stable for different values of parameters {FC, p -value}, giving rise to different orderings of the samples when different values for {FC, p -value} are used. Hence, we have conducted an analysis of the relationships between the different orderings of the disease samples and the quality of the approximation to the disease progression, using measures such as the increasing or decreasing trends of death rate, extracellular acidity, intracellular alkalinity, and a few others along the disease progression axis.

The result of the analysis indicates that the ordering that optimizes the following empirical function achieves the best approximation to the disease progression. Note that each set of values of {FC, p -value} gives rise to a distinct ordering of the samples and the associated $\{V_i\}$ values, where FC is selected from the range (1.1, 4.0) and p -value from (0.0005, 0.05), using 0.1 and 0.0001 as the increments for the former and the latter, respectively:

$$\max_{FC, p\text{-value}} \sum_{1 \leq i \leq |MCI \cup AD|} V_i$$

Subject to 1. the first $|MCI|$ samples in the current ordering contain at least $\gamma\%$ of MCI samples,
2. the last $(|MCI| + |AD|/2)$ samples contain at least $\gamma\%$ of AD samples,
where $|X|$ denotes the number of elements in set X , and γ is set to 55 based on our empirical analysis results.

The following provides a justification for why this list is useful for studies of the AD progression, along with its applications. We note that (1) the levels of both intracellular alkalization and extracellular acidification progressively go up with this approximation as shown in Figure 2A; and (2) the same is observed about the levels of $A\beta$ formation, Tau-fibril formation, oxidative stress, and neuronal apoptosis as shown in Figure 2A, which are consistent with previous studies [1,2], thus supporting the validity of this approximation scheme to disease progression.

For a given list of samples ordered as above, we partition it to four equal-sized sub-lists: bin1, bin2, bin3, and bin4 (possibly except for the last one), representing four periods of the disease evolution (Figures 2A).

4.2.5. ssGSEA Gene-Set Scores for Individual Samples

For gene-set enrichment analysis (GSEA) over individual samples, we employed the “`gsva()`” function in R, utilizing the “`ssgsea`” method, tailored specifically for GSEA on individual samples [56], which builds on the empirical cumulative distribution function (ECDF) of gene expression ranks. First, gene expression data are ranked across samples. Then, for each specified gene set, the ECDF is computed based on the ranks of the genes within and outside the set. This comparison allows the method to determine the extent to which genes within the gene set are enriched by the highly ranked genes. Positive enrichment scores signify an overrepresentation of the gene set among highly ranked genes, while negative scores indicate being enriched by lowly ranked genes.

4.2.6. Co-Expression Analyses

The following is used to estimate the co-expression level between a gene g and a gene set M in terms of their expression levels. Let PC_1 and PC_2 be the first and second principal components of the expression levels of gene set M . A linear regression model was constructed as below:

$$e(g) = \beta_1 PC_1 + \beta_2 PC_2 + \beta_0$$

where $\{\beta_i\}$ are parameters to be determined through minimizing $|\beta_0|$.

4.2.7. Predicting Target Genes or Pathways of Cis and Trans Regulation by lncRNA

A lncRNA can participate in the regulation of expressions of genes in its genomic neighborhood, referred to as *cis*-regulation, or genes distant from its genomic location, called *trans*-regulation [57,58]. Prediction of target genes regulated by a lncRNA via the *cis* mechanism relies on locational

relationships. Genes within a range of 100 kb upstream or downstream of a lncRNA are considered as *cis*-regulated target genes by the lncRNA if they are co-expressed with the lncRNA [59,60]. *Trans*-regulated genes by a lncRNA are determined based on sequence complementarity between the lncRNA and each target gene and being co-expressed between the two. Rsearch [61] is used here to predict *trans*-regulated genes by a lncRNA, where the prediction criteria also include the level of the free energy in the formed secondary structure between lncRNA and mRNA sequences (energy < -10) [62].

We further define the regulatory relationship between a lncRNA and a pathway as: a lncRNA is considered as a regulator of a pathway if an optimal regression model of the lncRNA against the principal components (PCs) of the pathway across all the relevant samples yields a high R^2 value and an F_{test} p -value < 0.05, and if the lncRNA is predicted to be a *cis* or *trans* regulator of any gene (or transcript) in the pathway, with a Spearman correlation coefficient $|r| > 0$ and a p -value < 0.05 and the lncRNA strongly correlated with the first PC of the pathway.

4.2.8. Construction of lncRNA-mRNA Interaction Network and Identification of Functional Modules

For a set of disease samples, we construct a lncRNA-mRNA interaction network, based on co-expressed lncRNA-mRNA pairs, using the method given [63], for all mRNAs of protein-coding genes. Specifically, for each pair of lncRNA and mRNA, we calculate the Pearson correlation coefficient (PCC) between the expression levels of the pair. All pairs with $|PCC| > 0$ constitute the interaction network, referred to as a *bi-color network*, following the definition in [64]. We have then applied the Markov clustering algorithm [65] to identify maximal subnetworks having dense intra-interactions of the network, referred to as *clusters*.

For all the identified clusters, we select those each consisting of at least 10% lncRNAs and at least three genes. Pathway enrichment is conducted over each set of gene clusters against three databases: KEGG [66], REACTOME [67], and GO Biological Process [56] using ClusterProfiler [68] and p -value < 0.05 as the threshold.

The enriched pathways are grouped into six categories: immune activities, metabolisms, stress-response, development-related, cell polarity, and neural functions. Here, we consider glial cells, such as astrocytes and microglia, as immune cells.

The idea in defining the mRNA-lncRNA interaction network and conducting the network-based clustering is to assign lncRNAs, particularly the novel ones, to pathways enriched by functionally well-annotated mRNAs, therefore giving some functional information to the lncRNAs.

4.2.9. The Discerning Power of lncRNAs in Distinguishing Control from AD Samples

Consider three sample sets: controls, bin_1 , and $\bigcup_{2 \leq i \leq 4} \text{bin}_i$, for normal, early, advanced AD samples, and $D = (\text{controls} \cup \bigcup_{1 \leq i \leq 4} \text{bin}_i)$, respectively. Let their sizes be s_1 , s_2 and s_3 . For the three sample sets, each of their element has a *label* 1, 2, or 3, if it belongs to the first, second or the third set, respectively. We aim to demonstrate that some lncRNA-containing pathways have the discerning power in separating the three sets.

For each pathway P_k considered, we use its first principal component across all the samples to represent the pathway. Define $\text{ave}(P_k, i)$ as the average expression level of all genes in P_k in sample $d_i \in D$. Sort all samples in D in the ascending order of $\text{ave}(P_k, i)$, and then partition the sorted list into three sub-lists D_1, D_2, D_3 with $|D_i| = s_i$, for $i \in [1,3]$. Define the *discerning* score as

$$S(D_1, D_2, D_3) = \sum_{d \in D_i, \text{label}(d) < i, i \in [1,3]} (i - \text{label}(d))$$

where $\text{label}(d)$ refers to the label of d as defined above. Consider the density distribution \mathcal{L} of $S(D_1, D_2, D_3)$ over all partitions (D_1, D_2, D_3) of D . We consider a P_k having discerning power if the p -value for its $S(D_1, D_2, D_3)$ in \mathcal{L} is < 0.05.

Supplementary Materials: The following supporting information can be downloaded at: <https://github.com/zhenyuh19/A-Model-for-the-Development-of-Alzheimer-s-Disease/tree/main>, Figure S1:

Summary of exon count and length for mRNA, old lncRNA, and new lncRNA; Figure S2: Evaluation of contribution of all four bin pathways to AD samples; Figure S3: Number of neuronal apoptosis pathways and neuron death promoted by lncRNA in the formation of Tau fibers; Figure S4: Statistics of lncRNA associated with pathologic hypotheses of AD; Figure S5: PCA contribution rate statistics of H⁺-producing enzyme genes in late-stage AD (Bin4); Table S1: Clinical information of ROSMAP prefrontal cortex samples; Table S2: The four bin pathways contributed to the six categories of pathway statistics in the top 20%; Table S3: Statistics of lncRNAs regulating microtubule depolymerization in cis and trans, along with mRNA enrichment results associated with lncRNAs regulating microtubule depolymerization; Table S4: Statistics of lncRNAs regulating Tau fibrillization in cis and trans, along with mRNA enrichment results associated with lncRNAs regulating Tau fibrillization; Table S5: Statistics of lncRNAs regulating amyloid formation in cis and trans, along with mRNA enrichment results associated with lncRNAs regulating amyloid formation; Table S6: Statistics of lncRNAs regulating hydrogen peroxide in cis and trans, along with mRNA enrichment results associated with lncRNAs regulating hydrogen peroxide and superoxide anion generation; Table S7: Statistics of lncRNAs regulating mitochondrial iron sulfur and copper transport in cis and trans, along with mRNA enrichment results associated with lncRNAs regulating mitochondrial iron sulfur; Table S8: Statistics of lncRNAs regulating extracellular copper in cis and trans, along with mRNA enrichment results associated with lncRNAs regulating mitochondrial iron sulfur; Table S9: Statistics of lncRNAs regulating synapse assembly in cis and trans, along with mRNA enrichment results associated with lncRNAs regulating synapse assembly; Table S10: Statistics of lncRNAs regulating cholesterol in cis and trans, along with mRNA enrichment results associated with lncRNAs regulating cholesterol; Table S11: Statistics of lncRNAs regulating H⁺-producing enzyme in cis and trans, along with mRNA enrichment results associated with lncRNAs regulating H⁺-producing enzyme; Table S12: PCA analysis of H⁺-producing enzymes reveals the ranking of each gene's contribution to PC1 explained variance ratio; Table S13: Statistics of lncRNAs regulating acid loading transporter in cis and trans, along with mRNA enrichment results associated with lncRNAs regulating acid loading transporter; Table S14: lncRNAs predicted as markers regulating each AD phenotype and cellular stress; Supplementary Data S1: New lncRNA Sequences and Genome Annotations.

Author Contributions: Conceptualization, Z.Y.H. and Y.X.; methodology, Z.Y.H. and Y.X.; software, Z.Y.H.; formal analysis, Z.Y.H.; investigation, Z.Y.H. and X.C.M.; data curation, Z.Y.H., Q.F.C., X.J.W and Z.A.; writing—original draft preparation, Z.Y.H.; writing—review and editing, Y.X.; visualization, Z.Y.H., Q.F.C., X.J.W; supervision, Y.X.; project administration, Y.X.; funding acquisition, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number T2350010 and The APC was funded by the Key Laboratory of Metabolism and Health of Guangdong Ordinary Colleges and Universities, grant number 2022KSYS007.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the reported results can be found at <https://github.com/zhenyuh19/A-Model-for-the-Development-of-Alzheimer-s-Disease/tree/main>.

Acknowledgments: The senior author thanks Southern University of Science and Technology for its start-up fund. We are grateful to the participants in the Religious Order Study, the Memory and Aging Project. Without their contributions, there will be not this project. Additionally, we extend our gratitude to Xiaojuan Wu for her assistance in data collection and result visualization.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- [1] L. Schwartz, S. Peres, M. Jolicoeur, and J. da Veiga Moreira, 'Cancer and Alzheimer's disease: intracellular pH scales the metabolic disorders', *Biogerontology*, vol. 21, no. 6, pp. 683–694, Dec. 2020, doi: 10.1007/s10522-020-09888-6.
- [2] Z.-G. Xiong, G. Pignataro, M. Li, S. Chang, and R. P. Simon, 'Acid-Sensing Ion Channels (ASICs) as Pharmacological Targets for Neurodegenerative Diseases', *Curr. Opin. Pharmacol.*, vol. 8, no. 1, pp. 25–32, Feb. 2008, doi: 10.1016/j.coph.2007.09.001.
- [3] P.-P. Guan, L.-L. Cao, Y. Yang, and P. Wang, 'Calcium Ions Aggravate Alzheimer's Disease Through the Aberrant Activation of Neuronal Networks, Leading to Synaptic and Cognitive Deficits', *Front. Mol.*

- Neurosci.*, vol. 14, 2021, Accessed: Sep. 03, 2022. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnmol.2021.757515>
- [4] V. M. Vitvitsky, S. K. Garg, R. F. Keep, R. L. Albin, and R. Banerjee, 'Na⁺ and K⁺ ion imbalances in Alzheimer's disease', *Biochim. Biophys. Acta*, vol. 1822, no. 11, pp. 1671–1681, Nov. 2012, doi: 10.1016/j.bbadis.2012.07.004.
 - [5] G. Verdile *et al.*, 'Inflammation and Oxidative Stress: The Molecular Connectivity between Insulin Resistance, Obesity, and Alzheimer's Disease', *Mediators Inflamm.*, vol. 2015, p. 105828, 2015, doi: 10.1155/2015/105828.
 - [6] L. Zuo *et al.*, 'Inflammaging and Oxidative Stress in Human Diseases: From Molecular Mechanisms to Novel Treatments', *Int. J. Mol. Sci.*, vol. 20, no. 18, p. 4472, Sep. 2019, doi: 10.3390/ijms20184472.
 - [7] D. W. Dickson, 'Apoptotic mechanisms in Alzheimer neurofibrillary degeneration: cause or effect?', *J. Clin. Invest.*, vol. 114, no. 1, pp. 23–27, Jul. 2004, doi: 10.1172/JCI200422317.
 - [8] H. Chi, H.-Y. Chang, and T.-K. Sang, 'Neuronal Cell Death Mechanisms in Major Neurodegenerative Diseases', *Int. J. Mol. Sci.*, vol. 19, no. 10, Art. no. 10, Oct. 2018, doi: 10.3390/ijms19103082.
 - [9] C. Haass and D. Selkoe, 'If amyloid drives Alzheimer disease, why have anti-amyloid therapies not yet slowed cognitive decline?', *PLOS Biol.*, vol. 20, no. 7, p. e3001694, Jul. 2022, doi: 10.1371/journal.pbio.3001694.
 - [10] T. Niikura, H. Tajima, and Y. Kita, 'Neuronal Cell Death in Alzheimer's Disease and a Neuroprotective Factor, Humanin', *Curr. Neuropharmacol.*, vol. 4, no. 2, pp. 139–147, Apr. 2006.
 - [11] X. Wang, W. Wang, L. Li, G. Perry, H. Lee, and X. Zhu, 'Oxidative Stress and Mitochondrial Dysfunction in Alzheimer's Disease', *Biochim. Biophys. Acta*, vol. 1842, no. 8, pp. 1240–1247, Aug. 2014, doi: 10.1016/j.bbadis.2013.10.015.
 - [12] T. Pohlkamp *et al.*, 'Endosomal Acidification by NHE6-depletion Corrects ApoE4-mediated Synaptic Impairments and Reduces Amyloid Plaque Load'. bioRxiv, Mar. 23, 2021. doi: 10.1101/2021.03.22.436385.
 - [13] E. Im *et al.*, 'Lysosomal dysfunction in Down Syndrome and Alzheimer mouse models is caused by selective v-ATPase inhibition by Tyr682 phosphorylated APP β CTF'. bioRxiv, Jan. 26, 2023. doi: 10.1101/2022.06.02.494546.
 - [14] A. Serrano-Pozo, M. P. Frosch, E. Masliah, and B. T. Hyman, 'Neuropathological Alterations in Alzheimer Disease', *Cold Spring Harb. Perspect. Med.*, vol. 1, no. 1, p. a006189, Sep. 2011, doi: 10.1101/cshperspect.a006189.
 - [15] E. Drummond and T. Wisniewski, 'Alzheimer's Disease: Experimental Models and Reality', *Acta Neuropathol. (Berl.)*, vol. 133, no. 2, pp. 155–175, Feb. 2017, doi: 10.1007/s00401-016-1662-x.
 - [16] E. O. Olufunmilayo and R. M. D. Holsinger, 'Roles of Non-Coding RNA in Alzheimer's Disease Pathophysiology', *Int. J. Mol. Sci.*, vol. 24, no. 15, p. 12498, Aug. 2023, doi: 10.3390/ijms241512498.
 - [17] Z. Lan, Y. Chen, J. Jin, Y. Xu, and X. Zhu, 'Long Non-coding RNA: Insight Into Mechanisms of Alzheimer's Disease', *Front. Mol. Neurosci.*, vol. 14, Jan. 2022, doi: 10.3389/fnmol.2021.821002.
 - [18] X. Shan, H. Tashiro, and C. G. Lin, 'The Identification and Characterization of Oxidized RNAs in Alzheimer's Disease', *J. Neurosci.*, vol. 23, no. 12, pp. 4913–4921, Jun. 2003, doi: 10.1523/JNEUROSCI.23-12-04913.2003.
 - [19] D. A. Bennett, A. S. Buchman, P. A. Boyle, L. L. Barnes, R. S. Wilson, and J. A. Schneider, 'Religious Orders Study and Rush Memory and Aging Project', *J. Alzheimers Dis. JAD*, vol. 64, no. Suppl 1, pp. S161–S189, 2018, doi: 10.3233/JAD-179939.
 - [20] A. Rijpmma, M. van der Graaf, O. Meulenbroek, M. G. M. Olde Rikkert, and A. Heerschap, 'Altered brain high-energy phosphate metabolism in mild Alzheimer's disease: A 3-dimensional 31P MR spectroscopic imaging study', *NeuroImage Clin.*, vol. 18, pp. 254–261, Feb. 2018, doi: 10.1016/j.nicl.2018.01.031.
 - [21] E. Lyros *et al.*, 'Normal brain aging and Alzheimer's disease are associated with lower cerebral pH: an in vivo histidine 1H-MR spectroscopy study', *Neurobiol. Aging*, vol. 87, pp. 60–69, Mar. 2020, doi: 10.1016/j.neurobiolaging.2019.11.012.
 - [22] Y. Decker *et al.*, 'Decreased pH in the aging brain and Alzheimer's disease', *Neurobiol. Aging*, vol. 101, pp. 40–49, May 2021, doi: 10.1016/j.neurobiolaging.2020.12.007.
 - [23] C.-X. Gong, I. Grundke-Iqbal, and K. Iqbal, 'Dephosphorylation of Alzheimer's disease abnormally phosphorylated tau by protein phosphatase-2A', *Neuroscience*, vol. 61, no. 4, pp. 765–772, Aug. 1994, doi: 10.1016/0306-4522(94)90400-6.
 - [24] S. Ciudad *et al.*, ' $\text{A}\beta(1-42)$ tetramer and octamer structures reveal edge conductivity pores as a mechanism for membrane damage', *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, Jun. 2020, doi: 10.1038/s41467-020-16566-1.
 - [25] A. A. Farooqui, 'Contribution of Dietary Carbohydrates in Induction of Oxidative Stress', in *Inflammation and Oxidative Stress in Neurological Disorders: Effect of Lifestyle, Genes, and Age*, A. A. Farooqui, Ed., Cham: Springer International Publishing, 2014, pp. 237–261. doi: 10.1007/978-3-319-04111-7_8.
 - [26] J. Harrow *et al.*, 'GENCODE: The reference human genome annotation for The ENCODE Project', *Genome Res.*, vol. 22, no. 9, pp. 1760–1774, Sep. 2012, doi: 10.1101/gr.135350.111.

- [27] A. Shumate, B. Wong, G. Pertea, and M. Pertea, 'Improved transcriptome assembly using a hybrid of long and short reads with StringTie', *PLOS Comput. Biol.*, vol. 18, no. 6, p. e1009730, Jun. 2022, doi: 10.1371/journal.pcbi.1009730.
- [28] C.-H. Tsai, T.-C. Lin, Y.-H. Chang, H.-K. Tsai, and J.-H. Huang, 'Identification and comparative analysis of long non-coding RNAs in the brain of fire ant queens in two different reproductive states', *BMC Genomics*, vol. 22, no. 5, p. 917, Apr. 2022, doi: 10.1186/s12864-022-08539-z.
- [29] Y. Liu *et al.*, 'Long non-coding RNA and circular RNA and coding RNA profiling of plasma exosomes of osteosarcoma by RNA seq', *Sci. Data*, vol. 10, no. 1, Art. no. 1, Jun. 2023, doi: 10.1038/s41597-023-02295-9.
- [30] 'Wilcoxon signed-rank test - Handbook of Biological Statistics'. Accessed: Mar. 13, 2024. [Online]. Available: <https://www.biostathandbook.com/wilcoxonsignedrank.html>
- [31] J.-Q. Ju, H.-L. Zhang, Y. Wang, L.-L. Hu, and S.-C. Sun, 'Kinesin KIFC3 is essential for microtubule stability and cytokinesis in oocyte meiosis', *Cell Commun. Signal.*, vol. 22, no. 1, p. 199, Mar. 2024, doi: 10.1186/s12964-024-01589-8.
- [32] S. Ozon, A. Guichet, O. Gavet, S. Roth, and A. Sobel, 'Drosophila Stathmin: A Microtubule-destabilizing Factor Involved in Nervous System Formation', *Mol. Biol. Cell*, vol. 13, no. 2, pp. 698–710, Feb. 2002, doi: 10.1091/mbc.01-07-0362.
- [33] T. Eckert, D. T.-V. Le, S. Link, L. Friedmann, and G. Woehlke, 'Spastin's Microtubule-Binding Properties and Comparison to Katanin', *PLoS ONE*, vol. 7, no. 12, p. e50161, Dec. 2012, doi: 10.1371/journal.pone.0050161.
- [34] C. M. Cowan and A. Mudher, 'Are Tau Aggregates Toxic or Protective in Tauopathies?', *Front. Neurol.*, vol. 4, p. 114, Aug. 2013, doi: 10.3389/fneur.2013.00114.
- [35] 'Introduction to Condensed Matter Chemistry - 1st Edition | Elsevier Shop'. Accessed: Apr. 16, 2024. [Online]. Available: <https://shop.elsevier.com/books/introduction-to-condensed-matter-chemistry/you/978-0-443-16140-7>
- [36] T. Rival *et al.*, 'Fenton chemistry and oxidative stress mediate the toxicity of the beta-amyloid peptide in a Drosophila model of Alzheimer's disease', *Eur. J. Neurosci.*, vol. 29, no. 7, pp. 1335–1347, Apr. 2009, doi: 10.1111/j.1460-9568.2009.06701.x.
- [37] R. Bauer, B. Brüne, and T. Schmid, 'Cholesterol metabolism in the regulation of inflammatory responses', *Front. Pharmacol.*, vol. 14, p. 1121819, Jan. 2023, doi: 10.3389/fphar.2023.1121819.
- [38] M. Trupp *et al.*, 'Beyond the genome (BTG) is a (PGDB) pathway genome database: HumanCyc', *Genome Biol.*, vol. 11, no. Suppl 1, p. O12, 2010, doi: 10.1186/gb-2010-11-s1-o12.
- [39] J. Gencel-Augusto, W. Wu, and T. G. Bivona, 'Long Non-Coding RNAs as Emerging Targets in Lung Cancer', *Cancers*, vol. 15, no. 12, p. 3135, Jun. 2023, doi: 10.3390/cancers15123135.
- [40] R. Li, H. Zhu, and Y. Luo, 'Understanding the Functions of Long Non-Coding RNAs through Their Higher-Order Structures', *Int. J. Mol. Sci.*, vol. 17, no. 5, p. 702, May 2016, doi: 10.3390/ijms17050702.
- [41] J. Luo, L. Qu, F. Gao, J. Lin, J. Liu, and A. Lin, 'LncRNAs: Architectural Scaffolds or More Potential Roles in Phase Separation', *Front. Genet.*, vol. 12, Mar. 2021, doi: 10.3389/fgene.2021.626234.
- [42] 'Extracellular pH modulates kinetics of the Na⁺,K⁺-ATPase', *Biochim. Biophys. Acta BBA - Biomembr.*, vol. 1509, no. 1–2, pp. 496–504, Dec. 2000, doi: 10.1016/S0005-2736(00)00356-4.
- [43] S. Janelidze *et al.*, 'Plasma β -amyloid in Alzheimer's disease and vascular disease', *Sci. Rep.*, vol. 6, p. 26801, May 2016, doi: 10.1038/srep26801.
- [44] T. K. Karikari *et al.*, 'Diagnostic performance and prediction of clinical progression of plasma phospho-tau181 in the Alzheimer's Disease Neuroimaging Initiative', *Mol. Psychiatry*, vol. 26, no. 2, pp. 429–442, Feb. 2021, doi: 10.1038/s41380-020-00923-z.
- [45] D. A. Bennett, J. A. Schneider, Z. Arvanitakis, and R. S. Wilson, 'OVERVIEW AND FINDINGS FROM THE RELIGIOUS ORDERS STUDY', *Curr. Alzheimer Res.*, vol. 9, no. 6, pp. 628–645, Jul. 2012.
- [46] R. Caspi *et al.*, 'The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases', *Nucleic Acids Res.*, vol. 40, no. Database issue, pp. D742–D753, Jan. 2012, doi: 10.1093/nar/gkr1014.
- [47] Y. Chen *et al.*, 'SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data', *GigaScience*, vol. 7, no. 1, Dec. 2017, doi: 10.1093/gigascience/gix120.
- [48] Y. Zhang, C. Park, C. Bennett, M. Thornton, and D. Kim, 'Rapid and accurate alignment of nucleotide conversion sequencing reads with HISAT-3N', *Genome Res.*, vol. 31, no. 7, pp. 1290–1295, Jan. 2021, doi: 10.1101/gr.275193.120.
- [49] Y.-J. Kang *et al.*, 'CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features', *Nucleic Acids Res.*, vol. 45, no. W1, pp. W12–W16, Jul. 2017, doi: 10.1093/nar/gkx428.
- [50] S. Han, Y. Liang, Y. Li, and W. Du, 'Long Noncoding RNA Identification: Comparing Machine Learning Based Tools for Long Noncoding Transcripts Discrimination', *BioMed Res. Int.*, vol. 2016, 2016, doi: 10.1155/2016/8496165.

- [51] L. Wang, H. J. Park, S. Dasari, S. Wang, J.-P. Kocher, and W. Li, 'CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model', *Nucleic Acids Res.*, vol. 41, no. 6, p. e74, Apr. 2013, doi: 10.1093/nar/gkt006.
- [52] A. Li, J. Zhang, and Z. Zhou, 'PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme', *BMC Bioinformatics*, vol. 15, no. 1, p. 311, Sep. 2014, doi: 10.1186/1471-2105-15-311.
- [53] X. Tong and S. Liu, 'CPPred: coding potential prediction based on the global description of RNA sequence', *Nucleic Acids Res.*, vol. 47, no. 8, pp. e43–e43, May 2019, doi: 10.1093/nar/gkz087.
- [54] M. I. Love, W. Huber, and S. Anders, 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol.*, vol. 15, no. 12, p. 550, Dec. 2014, doi: 10.1186/s13059-014-0550-8.
- [55] A. C. Frazee, G. Pertea, A. E. Jaffe, B. Langmead, S. L. Salzberg, and J. T. Leek, 'Ballgown bridges the gap between transcriptome assembly and expression analysis', *Nat. Biotechnol.*, vol. 33, no. 3, pp. 243–246, Mar. 2015, doi: 10.1038/nbt.3172.
- [56] A. Subramanian *et al.*, 'Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles', *Proc. Natl. Acad. Sci.*, vol. 102, no. 43, pp. 15545–15550, Oct. 2005, doi: 10.1073/pnas.0506580102.
- [57] C. P. Ponting, P. L. Oliver, and W. Reik, 'Evolution and functions of long noncoding RNAs', *Cell*, vol. 136, no. 4, pp. 629–641, Feb. 2009, doi: 10.1016/j.cell.2009.02.006.
- [58] A. E. Kornienko, P. M. Guenzl, D. P. Barlow, and F. M. Pauler, 'Gene regulation by the act of long non-coding RNA transcription', *BMC Biol.*, vol. 11, p. 59, May 2013, doi: 10.1186/1741-7007-11-59.
- [59] S. Liu *et al.*, 'Whole-Transcriptome RNA Sequencing Uncovers the Global Expression Changes and RNA Regulatory Networks in Duck Embryonic Myogenesis', *Int. J. Mol. Sci.*, vol. 24, no. 22, p. 16387, Nov. 2023, doi: 10.3390/ijms242216387.
- [60] X. Li *et al.*, 'Global identification and analysis revealed differentially expressed lncRNAs associated with meiosis and low fertility in autotetraploid rice', *BMC Plant Biol.*, vol. 20, p. 82, Feb. 2020, doi: 10.1186/s12870-020-2290-0.
- [61] A. Wenzel, E. Akbaşlı, and J. Gorodkin, 'Rlsearch: fast RNA–RNA interaction search using a simplified nearest-neighbor energy model', *Bioinformatics*, vol. 28, no. 21, pp. 2738–2746, Nov. 2012, doi: 10.1093/bioinformatics/bts519.
- [62] H. Sun *et al.*, 'Genome-wide transcriptional profiling and functional analysis of long noncoding RNAs and mRNAs in chicken macrophages associated with the infection of avian pathogenic E. coli', *BMC Vet. Res.*, vol. 20, 2024, doi: 10.1186/s12917-024-03890-7.
- [63] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, 'Permutation importance: a corrected feature importance measure', *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, May 2010, doi: 10.1093/bioinformatics/btq134.
- [64] X. Guo *et al.*, 'Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks', *Nucleic Acids Res.*, vol. 41, no. 2, p. e35, Jan. 2013, doi: 10.1093/nar/gks967.
- [65] A. J. Enright, S. Van Dongen, and C. A. Ouzounis, 'An efficient algorithm for large-scale detection of protein families', *Nucleic Acids Res.*, vol. 30, no. 7, pp. 1575–1584, Apr. 2002.
- [66] M. Kanehisa and S. Goto, 'KEGG: kyoto encyclopedia of genes and genomes', *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000, doi: 10.1093/nar/28.1.27.
- [67] 'Reactome'. Accessed: Dec. 29, 2019. [Online]. Available: <http://bmbdb.cbi.pku.edu.cn/Reactome.php>
- [68] T. Wu *et al.*, 'clusterProfiler 4.0: A universal enrichment tool for interpreting omics data', *The Innovation*, vol. 2, no. 3, p. 100141, Aug. 2021, doi: 10.1016/j.xinn.2021.100141.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.