Alzheimer-related genes show accelerated evolution

Anne Nitsche1 · Christian Arnold1,2,3 · Uwe Ueberham4 · Kristin Reiche5 · Jörg Fallmann1 · Jörg Hackermüller6,7 · Friedemann Horn2 · Peter F. Stadler1,2,8,9,10,11,12 · Thomas Arendt4

Received: 21 June 2017 / Revised: 27 January 2020 / Accepted: 3 February 2020 / Published online: 13 March 2020 © The Author(s) 2020. This article is published with open access

Abstract
Alzheimer’s disease (AD) is a neurodegenerative disorder of unknown cause with complex genetic and environmental traits. While AD is extremely prevalent in human elderly, it hardly occurs in non-primate mammals and even non-human-primates develop only an incomplete form of the disease. This specificity of AD to human clearly implies a phylogenetic aspect. Still, the evolutionary dimension of AD pathomechanism remains difficult to prove and has not been established so far. To analyze the evolutionary age and dynamics of AD-associated-genes, we established the AD-associated genome-wide RNA-profile comprising both protein-coding and non-protein-coding transcripts. We than applied a systematic analysis on the conservation of splice-sites as a measure of gene-structure based on multiple alignments across vertebrates of homologs of AD-associated-genes. Here, we show that nearly all AD-associated-genes are evolutionarily old and did not originate later in evolution than not-AD-associated-genes. However, the gene-structures of loci, that exhibit AD-associated changes in their expression, evolve faster than the genome at large. While protein-coding-loci exhibit an enhanced rate of small changes in gene structure, non-coding loci show even much larger changes. The accelerated evolution of AD-associated-genes indicates a more rapid functional adaptation of these genes. In particular AD-associated non-coding-genes play an important, as yet largely unexplored, role in AD. This phylogenetic trait indicates that recent adaptive evolution of human brain is causally involved in basic principles of neurodegeneration. It highlights the necessity for a paradigmatic change of our disease-concepts and to reconsider the appropriateness of current animal-models to develop disease-modifying strategies that can be translated to human.

Introduction
Alzheimer’s disease (AD) is an age-related chronic neurodegenerative disorder of unknown cause. Its etiology likely involves complex genetic and environmental traits [1]. Neuropathologically, it is characterized by extracellular neuritic plaques mainly consisting of aggregated Aβ-peptide derived from the much larger amyloid precursor protein (APP) and intracellular neurofibrillary tangles composed of fibrillar aggregates of the microtubule-associated protein tau in a hyperphosphorylated form. The discovery that autosomal dominant forms of AD are attributable to mutated genes coding for APP or one of the presenilins, proteins that contain the catalytic activity of gamma-secretase, releasing Aβ from APP, prompted continuous efforts to establish ‘transgenic mice models of AD’ for therapeutic research. Over the last 20 years, these transgenic mice have been state-of-the-art tools to develop preventive or therapeutic strategies against AD. Still, the tremendous success to ameliorate amyloid pathology and cognitive dysfunction in these mice models that have been observed in more than 300 reports could not be translated into effective therapies for AD patients [2–4]. While several attempts have been made to explain this apparent discrepancy [2], overall it highlights the necessity to reconsider our approaches to define the molecular pathology of AD and the

Supplementary information
The online version of this article (https://doi.org/10.1038/s41380-020-0680-1) contains supplementary material, which is available to authorized users.

These authors contributed equally: Christian Arnold, Uwe Ueberham, Kristin Reiche

Extended author information available on the last page of the article
appropriate use of model systems to validate therapeutic concepts.

AD is a human-specific disorder. While deposits of Aβ have been described in birds, fish, and various mammalian species, neurofibrillary tangles are found almost exclusively in humans, and even non-human primate only develop an incomplete form of the disease [5–7]. This exclusivity of AD to human brain clearly implies a phylogenetic aspect of the disease and most likely indicates that adaptive changes of cerebral structure and function that have occurred during human evolution may have rendered the human brain sensitive to AD [8]. Clear cut phylogenetic traits of the AD pathomechanism might have far reaching consequences with respect to our approaches of disease prevention and therapy including defining appropriate animal model systems. Still, this evolutionary dimension remains difficult to prove and has not been established unequivocally so far.

Major phenotypic changes that have occurred in the course of human evolution, especially those between humans and chimpanzees, are suggested to mainly result from an increase in gene expression and are, thus, reflected at the transcriptomic level [9–11]. Genes whose expression have increased in human brain are mainly related to growth and differentiation [12] and frequently are involved in transcriptional regulation and RNA processing [9, 10]. While major evolutionary changes might have occurred at the transcriptomic level, they appear to be particularly pronounced for non-coding RNAs (ncRNAs). As shown by analyses of sequenced genomes of a large variety of species, the relative amount of non-coding sequence increases consistently with complexity [13]. Thus, ncRNAs, most likely constitute a critical layer of gene regulation in complex organisms that have expanded during evolution [14]. They have been conceptualized as a “complexity multipliers” in the human brain, allowing the ~30 000 protein coding genes to code for ~10^6 synapses [15].

To test the hypothesis that brain evolution critically contributes towards the AD pathomechanism, here, we established the AD-associated genome-wide RNA profile comprising both protein-coding (cRNA) and non-protein-coding (ncRNA) transcripts and applied a systematic analysis on the conservation of splice sites as a measure of the evolution of gene structure.

**Methods**

**Patients and healthy controls**

Brain tissue of 19 AD patients and 21 healthy controls dying without any history of neurological or psychiatric illness was used. The diagnosis of AD was made on the basis of both clinical and neuropathological evidence according to the criteria of the International Working Group (IWG) for New Research Criteria for the diagnosis of AD in the revision of 2014 (IWG-2) [16], the NIA-AA diagnostic criteria in the revision of 2011 [17] and the NIA-AA guidelines for the neuropathological assessment of AD [18, 19]. Only cases with typical AD according to IWG-2 criteria were included. All cases underwent neuropsychological assessment within the last 6 months prior to their death. Clinical Dementia Rating (CDR) scale scoring was based on neuropsychological testing (CERAD) [20], MMSE [21] and rating scales [22]. All cases were neuropathologically assessed for NFT stage, Aβ/amyloid plaque score and for neuritic plaque score according to [18]. NFTs and Aβ/amyloid plaques were detected by immunocytochemical labeling of phospho-tau (anti-human PHF-tau monoclonal antibody AT8; Thermo Scientific) and Aβ (beta amyloid monoclonal antibody, 6E10; BioLegend), respectively. Case recruitment, autopsy and data handling have been performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments as well as with the convention of the Council of Europe on Human Rights and Biomedicine and had been approved by the responsible Ethics Committee of Leipzig University (GZ 01GI9999-01GI0299; Approval # 282–02). Informed consent was obtained from all subjects or their legal representatives.

**Genome-wide RNA profile and analysis of splice site conservation**

The AD-associated genome-wide RNA profile comprising both protein-coding (cRNA) and non-protein-coding (ncRNA) transcripts was established combining whole genome tiling arrays with a custom array approach. To this end, we designed a custom array comprising 931 898 probes derived from Agilent’s Whole Human Genome Oligo array, long non-coding RNA (lncRNA) probes extracted from public databases, computationally predicted loci of structured RNAs, and lncRNA probes experimentally identified by transcriptome-wide expression variation studies based on the Affymetrix Human Tiling 1.0 array comparing AD patients with control samples (Supplementary methods and results; Supplementary Tables S1, S2). This custom array allows us to not only rely on available annotation for transcripts, but extends the search space by computationally predicted loci and expression studies. Applying this custom array to 17 AD patients and 19 control samples, we identified a differential expression of 154 multi-exonic cRNAs with a total of 4 162 splice sites and 141 multi-exonic lncRNAs with a total of 1 297 splice sites (Supplementary methods and results, Section 2.2). In Fig. 1 we show a heatmap representation of differentially expressed probes highlighting the commonalities within and
differences between the AD and control patient groups. While Supplementary Table S3 provides an overlap of identified AD-associated genes with known AD-associated genes, differentially expressed coding and non-coding genes are listed in Supplementary Tables S4, S5.

To establish a measure of the evolution of gene structure, we applied a systematic analysis on the conservation of splice sites based on multiple alignments [23] across 18 vertebrates (Supplementary Tables S6, S7) of homologs of AD-associated protein-coding and non-coding genes (Supplementary methods and results, Section 1.7).

**Results**

In order to compare the conservation of genes at a structural level we consider two levels of resolution. We can ask whether the gene is present or absent in another species as an entity, i.e., without considering details of the intron-exon structure. This amounts to pinpointing its evolutionary origin. More stringently we may ask whether the exact intron-exon structure, and hence the human layout, is present in a non-human primate. Genes that are present but have changed their structure likely have undergone more dramatic functional changes than genes whose structure has been exactly preserved. Protein-coding genes and non-protein-coding genes were independently investigated for their conservation (Supplementary Fig. S1).

Nearly all AD-associated protein-coding genes are evolutionarily old (Fig. 2c). There were no differences in conservation rate when comparing presence of AD-associated and all protein-coding genes. This means that AD-associated protein-coding genes did not originate later in evolution than other protein-coding genes. In line with previous reports [14], IncRNAs are much less well conserved and many have emerged in the course of mammalian evolution. The fraction of conserved IncRNAs thus decreases rapidly with evolutionary time (Fig. 2a, b; Supplementary Figs. S2, S3). As for protein-coding sequences (Fig. 2c, d) we did not observe a significantly younger origin of AD-associated genes.
While there is no recognizable difference in the evolutionary age of origin between AD-associated genes compared to the transcriptome as a whole (Fig. 2c), we observed significant, albeit more subtle differences in the evolution of AD-associated and lncRNAs in general, concerning the changes in gene structure. The conservation rate of AD-associated non-coding genes decreases significantly \( (p < 0.05) \) below background level (Fig. 2a, b) not only for distantly related mammals but even primates, when complete conservation of gene structure is considered. In other words, the fraction of transcripts that have the entirety of their splice sites conserved is smaller amongst AD-associated ncRNAs than amongst non-coding genes at large. AD-associated ncRNAs hence show an accelerated evolution of their gene structure. This is indicative of a more rapid functional adaptation of AD-associated non-coding genes.

Although protein-coding genes are much better conserved than lncRNAs we observed the same increase of splice site turnover in AD. In fact, the relative effect is even stronger compared to non-protein-coding loci (\( \sim 30-40\% \) vs. \( \sim 5-15\% \) difference, shown as red lines in Fig. 2b, d, respectively). Since the same fraction of transcripts is present, while the conservation rate decreases with a more stringent conservation level of gene structure, we conclude that splice sites are systematically less conserved in human AD-associated regions compared to the typical behavior of the transcriptome.

While protein-coding loci exhibit an enhanced rate of small changes in their gene structure, we observe large changes in lncRNAs, again with a significantly enhanced rate in the AD-associated ncRNAs. The observed trends are robust as controls against possible confounding effects, e.g., of alignment quality demonstrate (Supplementary Figs. S4, S5).
This suggests that in particular AD-associated non-coding genes play an important, as yet largely unexplored, role in the AD pathomechanisms.

**Discussion**

Genome-wide studies that systematically analyze the evolutionary age of protein-coding and non-protein-coding AD-associated genes have not been performed previously. While major evolutionary changes might have occurred at the transcriptomic level, they appear to be particularly pronounced for IncRNAs [23, 24]. As shown by analyses of sequenced genomes of a large variety of species, the relative amount of non-coding sequence increases consistently with complexity [13]. Thus, IncRNAs, most likely constitute a critical layer of gene regulation in complex organisms that has expanded during evolution [14]. However, the evolutionary histories of IncRNAs have been notoriously hard to study due to their usually low level of sequence conservation. This not only hampers comprehensive homology-based annotation efforts but also makes it nearly impossible to obtain the high fidelity sequence alignments that are required for in depth studies into their evolution. Alternatively, the conservation of gene structure and particularly the conservation of splice sites may also be used to establish homology of IncRNAs [23]. Splice sites therefore leave “phylogenetic footprints”, and conserved patterns of splice sites may be used to predict novel transcripts from multiple genome alignments [25, 26]. Although IncRNAs are clearly ancient components of vertebrate genomes, they exhibit a rapid turnover of their intron/exon structures [23] that may be indicative of functional adaptation.

While the disease-relevance of IncRNAs is increasingly recognized, previous systematic gene expression profiling studies in AD nevertheless focused predominantly on protein-coding genes. Consequently, so far, only a few individual AD-associated ncRNAs have been identified and functionally characterized [27].

We have shown here that gene structures of AD-associated loci evolve faster than the genome at large, while there is no evidence that AD-associated genes originated particularly late in evolution. In order to capture the evolution of IncRNAs, we focused on gene structure, i.e., the conservation of splice sites because this approach makes it possible to separate the evolution of the transcripts from other selective constraints such as regulatory DNA elements that may affect sequence conservation [23]. Changes in gene structure can be expected to have in general larger functional effects than point mutations. The enhanced rate of gene structure evolution in AD-related genes hints a relation of AD to recent adaptive evolution, presumably in relation to the rapid evolution of the human brain.

Importantly, replacing the background set by only genes expressed in brain did not affect the conclusions (Supplementary methods and results, Section 2.3).

The most pronounced cortical expansion during hominid evolution occurs in the association cortex [28], i.e., brain areas that show most significant differences in gene expression between the human and non-human primate brain [10, 29], and are affected in AD most early and most constantly [30]. Evolutionary expansion of the neocortex, and in particular phylogenetic shaping of association areas, is associated with a developmental deceleration and an extended period of high neuronal plasticity into adulthood [12]. The presence of these neurons which remain structurally immature throughout their lifespans might provide the prerequisite both for the human adaption to the “cognitive niche” and for a high vulnerability towards factors that lead to the development of AD [31].

Our data support the concept that neurodegeneration in AD is a result of the evolutionary legacies that have occurred during the course of evolution of the human brain. This might explain why the ability to develop AD-type pathology has emerged only in lower primates and further complemented along the line leading to monkeys, great apes and humans [5], and why early symptoms of AD typically affect mental abilities that have been acquired only during very recent hominid evolution and as such are specific to human [32].

Our results also suggest that the mode of action of a number of lifestyle factors that have been identified as disease modifying factors such as physical activity, body fat, systolic blood pressure, stress, alcohol and tobacco consumption might be viewed against the background of genetic and environmental conditions that have resulted in the evolutionary emergence of humans in their current biological and socio-cultural form. In addition, genetic risk factors such as for instance ApoE, one of the strongest genes contributing to AD, should be considered against the background of adaptive mechanism occurring during hominid brain evolution [33].

This phylogenetic trait of AD highlights the necessity for a paradigmatic change of our concepts of the disease. Non-primate mammals, including rodents, are not very likely reliable models able to adequately mimic critical pathogenetic events of AD and to identify potential therapeutic targets. We thus need to reconsider our approaches to define the molecular pathology of AD and the appropriateness of current animal model systems to develop disease-modifying strategies.

**Data availability**

Tiling Array and Custom Array data are available at Gene Expression Omnibus (GSE138261).
Acknowledgements The study was supported by grants from the DFG (SPP1738: AR 200/17-1; HA 5857/2-1; STA 850/19-1) to P.F.S., T.A. and J.H. Open access funding provided by Projekt DEAL.

Author contributions T.A., U.U., K.R., C.A., F.H. and J.H. designed and performed the tiling array and custom array experiments. P.F.S., K.R., C.A., J.H. and J.F. performed bioinformatics analysis of the expression data. P.F.S. and A.N. developed and performed the computational method to assess the conservation of splice sites.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer’s disease. Cell. 2013;153:707–20.

2. Zahs KR, Ashe KH. ‘Too much good news’ - are Alzheimer mouse models trying to tell us how to prevent, not cure, Alzheimer’s disease? Trends Neurosci. 2010;33:381–9.

3. Gandy S, DeKosky ST. Toward the treatment and prevention of Alzheimer’s disease: rational strategies and recent progress. Annu Rev Med. 2013;64:367–83.

4. Li C, Ebrahimi A, Schluesser H. Drug pipeline in neurodegeneration based on transgenic mice models of Alzheimer’s disease. Ageing Res Rev. 2013;12:116–40.

5. Perez SE, Raghanti MA, Hof PR, Kramer L, Ikonomovic MD, Lacor PN, et al. Alzheimer’s disease pathology in the neocortex and hippocampus of the western lowland gorilla (Gorilla gorilla gorilla). J Comp Neurol. 2013;521:4318–38.

6. Finch CE, Austad SN. Commentary: is Alzheimer’s disease uniquely human? Neurobiol Aging. 2015;36:553–5.

7. Larsen PA, Luiz MW, Hunnicutt KE, Mihovilovic M, Saunders AM, Yoder AD, et al. The Alu neurodegeneration hypothesis: A primate-specific mechanism for neuronal transcription noise, mitochondrial dysfunction, and manifestation of neurodegenerative disease. Alzheimer’s Dement J Alzheimer’s Assoc. 2017;13:828–38.

8. Rapoport SI. Hypothesis: Alzheimer’s disease is a phylagenetic disease. Med Hypotheses. 1989;29:147–50.

9. Uddin M, Wildman DE, Liu G, Xu W, Johnson RM, Hof PR, et al. Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. Proc Natl Acad Sci USA. 2004;101:2957–62.

10. Gaceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, et al. Elevated gene expression levels distinguish human from non-human primate brains. Proc Natl Acad Sci USA. 2003;100:13030–5.

11. Fukuda K, Ichiyanagi K, Yamada Y, Go Y, Udono T, Wada S, et al. Regional DNA methylation differences between humans and chimpanzees are associated with genetic changes, transcriptional divergence and disease genes. J Hum Genet. 2013;58:446–54.

12. Somel M, Franz H, Yan Z, Lorenc A, Guo S, Giger T, et al. Transcriptional neoteny in the human brain. Proc Natl Acad Sci USA. 2009;106:5743–8.

13. Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays. 2007;29:288–99.

14. Mattick JS. The functional genomics of noncoding RNA. Science. 2005;309:1527–8.

15. Nelson PT, Keller JN. RNA in brain disease: no longer just “the messenger in the middle”. J Neuropathol Exp Neurol. 2007;66:461–8.

16. Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for Alzheimer’s disease: the IWG-2 criteria. Lancet Neurol. 2014;13:614–29.

17. McKhann GM, Knopman DS, Cherntow K, Hyman BT, Jack CR, JR, et al. The diagnosis of dementia due to Alzheimer’s disease: recommendations from the National Institute on Aging-Alzheimer’s Association workgroups on diagnostic guidelines for Alzheimer’s disease. Alzheimer’s Dement J Alzheimer’s Assoc. 2011;7:263–9.

18. Montine TJ, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Dickson DW, et al. National Institute on Aging-Alzheimer’s Association guidelines for the neuropathologic assessment of Alzheimer’s disease: a practical approach. Acta Neuropathol 2012;123:1–11.

19. Hyman BT, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Carrillo MC, et al. National Institute on Aging-Alzheimer’s Association guidelines for the neuropathologic assessment of Alzheimer’s disease. Alzheimer’s Dement J Alzheimer’s Assoc. 2012;8:1–13.

20. Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer’s Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer’s disease. Neurology. 1989;39:1159–65.

21. Folstein MF, Folstein SE, McHugh PR. ‘Mini-mental state’: A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.

22. Reisberg B, Ferris SH, Leon MJde, Crook T. The Global Deterioration Scale for assessment of primary degenerative dementia. Am J Psychiatry. 1982;139:1136–9.

23. Nitsche A, Rose D, Fasold M, Reiche K, Stadler PF. Comparison of splice sites reveals that long noncoding RNAs are evolutionarily well conserved. RNA. 2015;21:801–9.

24. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Stadler PF. Computational discovery of human coding and non-coding transcripts with conserved splice sites. Bioinformatics. 2011;27:1894–900.

25. Zhang Z. Long non-coding RNAs in Alzheimer’s disease. Curr Top medicinal Chem. 2016;16:511–9.
28. Stephan H. Evolutionary trends in limbic structures. Neurosci Biobehav Rev 1983;7:367–74.
29. Caceres M, Suwyn C, Maddox M, Thomas JW, Preuss TM. Increased cortical expression of two synaptogenic thrombospondins in human brain evolution. Cereb Cortex. 2007;17:2312–21.
30. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. Acta Neuropathol. 1991;82:239–59.
31. Arendt T. Alzheimer’s disease as a loss of differentiation control in a subset of neurons that retain immature features in the adult brain. Neurobiol Aging. 2000;21:783–96.
32. Matteson MA, Linton AD, Barnes SI. Cognitive developmental approach to dementia. Image J Nurs Scholarship. 1996;28:233–40.
33. Finch CE, Sapolsky RM. The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms. Neurobiol aging. 1999;20:407–28.

Affiliations

Anne Nitsche1 · Christian Arnold1,2,3 · Uwe Ueberham4 · Kristin Reiche5 · Jörg Fallmann1 · Jörg Hackermüller6,7 · Friedemann Horn2 · Peter F. Stadler1,2,8,9,10,11,12 · Thomas Arendt4

1 Bioinformatics Group, Department of Computer Science, University Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany
2 Department of Diagnostics, Fraunhofer Institute for Cell Therapy and Immunology – IZI, Perlickstraße 1, D-04103 Leipzig, Germany
3 Computational EvoDevo, Department of Computer Science, University Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany
4 Paul Flechsig Institute of Brain Research, Liebigstraße 19, 04103 Leipzig, Germany
5 Bioinformatics Unit, Department of Diagnostics, Fraunhofer Institute for Cell Therapy and Immunology – IZI, Perlickstraße 1, D-04103 Leipzig, Germany
6 Young Investigators Group Bioinformatics and Transcriptomics, Department Molecular Systems Biology, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany
7 Department of Computer Science, University Leipzig, Leipzig, Germany
8 Interdisciplinary Center for Bioinformatics, University Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany
9 Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103 Leipzig, Germany
10 Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria
11 Center for non-coding RNA in Technology and Health, University of Copenhagen, Grønnegårdsvej 3, DK-1870 Frederiksberg C, Denmark
12 Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA