Review and Progress Open Access

Genomic Prediction and its Association with the Development of Dementia disease in the Elderly

Li Xiaojun, Zhang Shuiji
Biotechnology Research Center, Cuixi Academy of Biotechnology, Zhuji, 311800, China
Corresponding author email: jessi.j.zhang@foxmail.com
Computational Molecular Biology, 2024, Vol.14, No.1 doi: 10.5376/cmb.2024.14.0001
Received: 29 Dec., 2023
Accepted: 30 Dec., 2023
Published: 04 Jan., 2024

Preferred citation for this article:
Li X.J., and Zhang S.J., 2024, Genomic prediction and its association with the development of dementia disease in the elderly, Computational Molecular Biology, 14(1): 1-8 (doi: 10.5376/cmb.2024.14.0001)

Abstract Dementia is a severe neurological disorder involving complex interactions between various genetic and environmental factors. This paper explores the association between genomic prediction and the development of dementia in the elderly. Through a systematic review of existing research, the study delves into genomics, the genetic basis of dementia, and the etiology related to the genome. The research further examines the methods and applications of genomic prediction, focusing on the use of polygenic risk scores and machine learning algorithms in dementia studies. Through case analyses of large-scale genomic studies, key genes associated with dementia, such as Alzheimer's disease, are revealed. Additionally, the paper thoroughly analyzes the major findings of existing research, emphasizing the filling of knowledge gaps and the provision of new insights. Finally, the paper discusses the challenges faced by genomic prediction, including methodological difficulties, challenges in data interpretation, ethical and privacy concerns, and more. Looking ahead to future research directions, the paper highlights the establishment of personalized genomic prediction models, the application of new technologies, and the potential value of genomic prediction in early diagnosis and prevention of dementia.
Keywords Elderly dementia disease; Genomic prediction; Genetics; Polygenic risk scores; Machine learning algorithms

Alzheimer's disease is a group of diseases mainly characterized by cognitive dysfunction, including Alzheimer's disease, vascular dementia, dancing disease, and frontotemporal dementia (Wu et al., 2021). According to statistics, Alzheimer's disease is an increasingly serious health problem among the elderly population worldwide, causing heavy burdens on patients and their families. With the trend of aging population, the incidence of Alzheimer's disease is on the rise, becoming an urgent problem to be solved in the medical field.

The definition of Alzheimer's disease not only covers cognitive decline, but also includes the impact on individuals' daily living abilities. Epidemiological data of this condition show that its incidence is closely related to age and there are gender differences. According to the report of the World Health Organization, Alzheimer's disease has become a major health challenge for the elderly population worldwide (Fagundes et al., 2011). It is estimated that by 2050, the number of patients with Alzheimer's disease will exceed 200 million, posing a serious threat to the sustainability of the global health system (Nichols et al., 2022).

Although in the past few decades, scientists have made significant progress in the etiology and pathophysiology of Alzheimer's disease (Simonetti et al., 2020), a radical cure has not yet been found. Therefore, more and more research is focused on understanding the genetic basis of Alzheimer's disease in order to intervene and treat it earlier.

Due to the rapid development of genomics technology, genomic prediction has gradually become a popular direction in medical research. This method analyzes variations in an individual's genome to predict their risk of developing a specific disease. In the field of Alzheimer's disease, genomic prediction provides a new way to understand the role of genetic factors in the development of the disease (Oriol et al., 2019).

Genomic prediction is a method based on genetic variation information to estimate individual susceptibility to a certain disease. Past studies have shown that Alzheimer's disease has a significant genetic predisposition, so using genomic prediction tools to explore its genetic basis has become the focus of scientists' attention.

With the continuous progress of high-throughput sequencing technology, researchers can more comprehensively interpret individual genomes and identify genetic variations associated with Alzheimer's disease. Genomic prediction has achieved some encouraging results in other disease areas, such as breast cancer and diabetes, providing reference and inspiration for the study of Alzheimer's disease.

The purpose of this review is to systematically summarize the association studies between genomic prediction and the development of Alzheimer's disease, deeply analyze the main findings of existing literature, and explore future directions in this research field. This review will review the methods and applications of genomic prediction, focusing on the identified genetic variations associated with Alzheimer's disease and how these variations affect the development of the disease.

Through this review, we aim to provide a comprehensive understanding of the genetics of Alzheimer's disease to the scientific community, and explore the potential applications of genomic prediction in early diagnosis, risk assessment, and personalized treatment. Ultimately, we hope to provide new theoretical support and research directions for the prevention and treatment of Alzheimer's disease.

1 Genomics and The Concept of Alzheimer's Disease

1.1 Definition of genomics

Genomics, as a discipline that studies the entire genome, is an important branch in the field of biology (McGuire et al., 2020). The genome is the collection of all genes within an organism, and genes are DNA fragments that carry genetic information and are responsible for encoding proteins or regulating the expression of other genes.

With the rapid development of technology, especially the application of high-throughput sequencing technology, it has become possible to fully understand the structure and function of the genome. The human genome consists of approximately 300 million base pairs and contains more than 20,000 genes. These genes carry instructions necessary for building and maintaining life, and their normal function is crucial for individual health.

One of the main goals of genomics is to understand the arrangement, function, and interrelationships of genes in the genome. Whole-genome sequencing determines the location of genes, the proteins they encode, and regulatory elements that control gene expression. This provides researchers with an opportunity to delve deeper into the genome to reveal the molecular basis of various physiological and pathological processes in the body (Hu et al., 2021).

In the study of Alzheimer's disease, the application of genomics can provide a more comprehensive understanding of patients' genetic information, especially those genetic variations associated with the development of Alzheimer's disease. By comparing the genomes of patients with those of healthy controls, scientists can identify specific genes, genomic regions, and variant types associated with Alzheimer's disease risk.

These genomic findings provide powerful tools for Alzheimer's disease research, enabling researchers to delve deeper into the genetic basis, biological mechanisms, and potential therapeutic targets.

1.2 Genetic basis of Alzheimer's disease

The genetic basis of Alzheimer's disease is complex and diverse, involving the interaction of multiple genes and environmental factors (Santos et al., 2020). The study of genetic factors helps to identify specific genes associated with Alzheimer's disease, providing important information for genomic prediction.

Studies have found that some forms of Alzheimer's disease have a family history, indicating the role of genetics in the development of the disease. Although most cases are considered to be polygenic, there are also some forms of familial Alzheimer's disease that are associated with single-gene mutations, such as early-onset familial Alzheimer's disease.

Alzheimer's disease, the most common form of dementia, has received extensive attention in terms of its genetic basis (Knopman et al., 2021). The genetic basis mainly involves genes related to amyloid precursor protein (APP), steroid hormone receptors (APOE), and others. Mutations or variations in these genes may lead to abnormal aggregation of amyloid and neuronal damage, ultimately leading to cognitive decline.

1.3 Genome-related Alzheimer's disease gene

In past studies, several genes have been identified as associated with the development of Alzheimer's disease. With technological advancements, large-scale genome-wide association studies (GWAS) have identified many new genes associated with Alzheimer's disease risk (Figure 1) (Bellenguez et al., 2022).

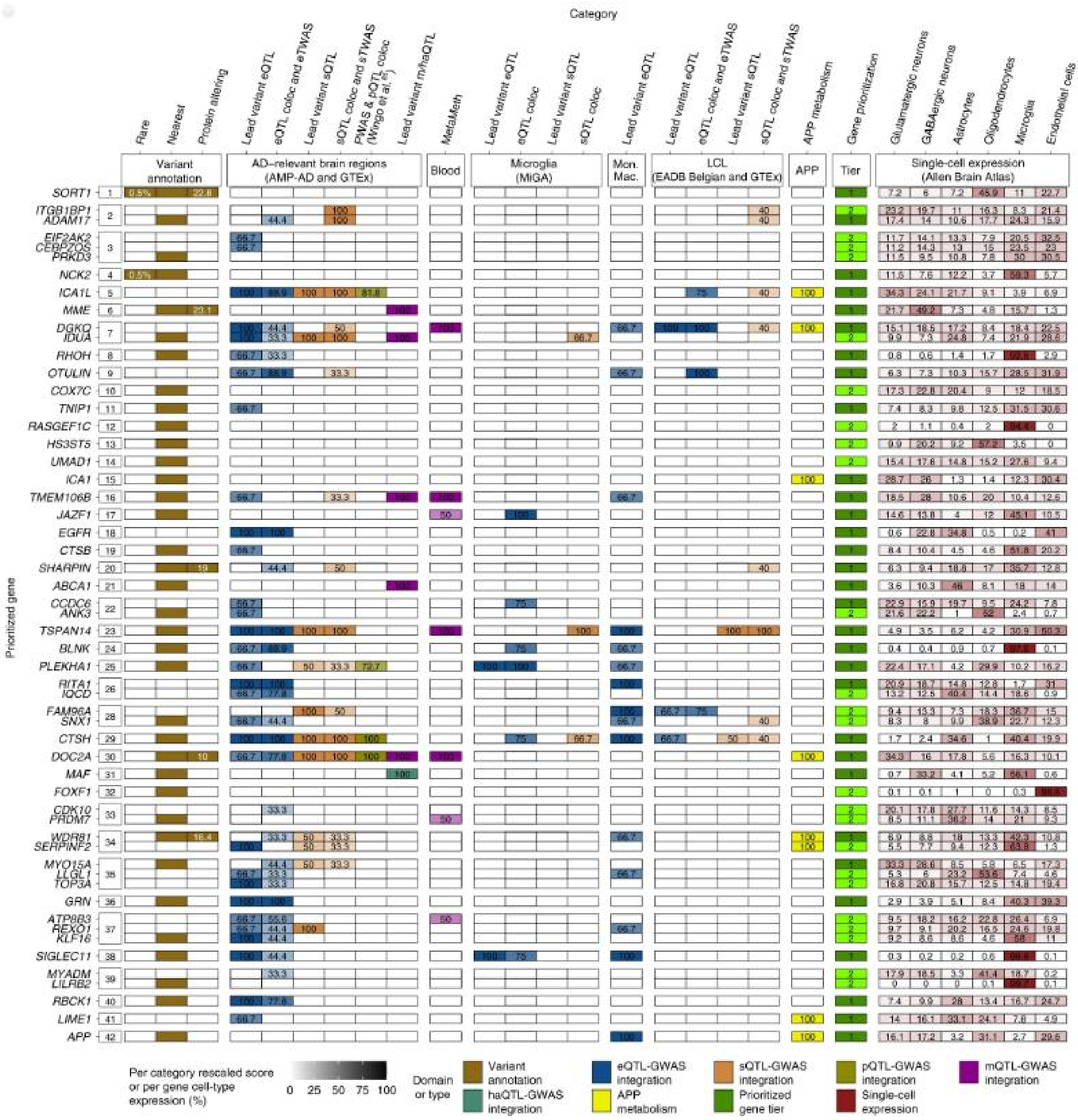

Figure 1 Gene prioritization of related dementias (ADD) (Bellenguez et al., 2022)

The APOE (apolipoprotein E) gene is one of the most closely associated genes with Alzheimer's disease. The ε4 allele of this gene is widely recognized as a major risk factor for Alzheimer's disease. Its mutant form is associated with increased amyloid deposition in the brain and neuronal damage, increasing the risk of developing Alzheimer's disease.

The CLU (Clusterin) gene encodes the clusterin protein, which is involved in amyloid deposition and clearance in the brain. Multiple studies have found that variations in the CLU gene are associated with an increased risk of developing Alzheimer's disease. Clusterin is thought to play an important role in anti-inflammatory and neuroprotective functions in the brain, and its abnormalities can lead to inflammatory responses and neuronal death.

The PICALM (Phosphatidylinositol Binding Clathrin Assembly Protein) gene encodes a protein that is involved in intracellular organelle transport in neurons. Variations in the PICALM gene are associated with an increased risk of developing Alzheimer's disease, possibly due to its role in promoting amyloid clearance and regulating neuronal survival (Ando et al., 2022).

Variations in the BIN1 (Bridging Integrator 1) gene are strongly associated with the risk of developing Alzheimer's disease (Gao et al., 2021). The BIN1 protein is involved in the regulation of membrane morphology, and its abnormalities can lead to amyloid deposition and neuronal damage.

Some variations in the SORL1 (Sortilin-Related Receptor 1) gene have been found to be associated with an increased genetic risk of Alzheimer's disease. The SORL1 protein plays a role in amyloid clearance and trafficking, and its abnormalities can lead to abnormal amyloid aggregation.

These genes are involved in multiple biological processes, and their mutations or variations can affect the pathogenesis of Alzheimer's disease through various pathways, including abnormal amyloid deposition, neuronal damage, and inflammatory responses. These findings provide a more comprehensive understanding of the pathogenesis of Alzheimer's disease and provide clues for exploring potential therapeutic targets.

1.4 The impact of the interaction between multiple genes and the environment

The development of Alzheimer's disease is not only influenced by single genes, but also by the complex interactions between multiple genes and the environment. The introduction of polygenic risk scores (PRS) allows researchers to comprehensively consider the contribution of multiple genetic variants to disease risk. Additionally, environmental factors such as lifestyle, education level, psychosocial factors, etc. have also been found to be associated with the risk of Alzheimer's disease, and their interactions with genes further increase the complexity of the research.

Researchers are working hard to reveal the interactions between these multiple factors to more comprehensively and accurately assess individual risk of Alzheimer's disease. Understanding these complex genetic and environmental interactions is crucial for developing prevention strategies and personalized treatment plans.

2 Methods and Applications of Genome Prediction

2.1 Basic principles of genome prediction

Genome prediction is a method that predicts an individual's susceptibility to a certain disease by analyzing the genetic variations in their genome. The basic principle of this method is to establish a model that associates known genetic variations associated with the disease with disease risk, and then use this model to analyze the individual's genome data and estimate their likelihood of developing the disease.

In genome prediction, commonly used methods include Polygenic Risk Scores (PRS) and machine learning algorithms (Lambert et al., 2019). PRS calculates a score by summing the risk weights of multiple genetic variants in an individual's genome, reflecting their overall genetic risk for a certain disease. Machine learning algorithms learn the genetic characteristics of the disease from a large amount of genome data, and then predict the risk for new individuals.

2.2 The application of genome prediction in Alzheimer's disease research

As a cutting-edge technology, genome prediction has already shown great potential in Alzheimer's disease research. Its main areas of application include risk assessment, early diagnosis, and a deeper understanding of the genetics of Alzheimer's disease.

By analyzing individual genome data, researchers can calculate the individual's likelihood of developing the disease, This enables more accurate personalized risk assessment. This is important for identifying high-risk populations, optimizing resource allocation, and developing personalized prevention strategies. For example, some researchers have successfully identified high-risk groups for Alzheimer's disease by constructing Polygenic Risk Scores (PRS) (Clark et al., 2022), providing strong support for personalized health management.

By analyzing genome data, researchers can identify genetic markers associated with early lesions, providing the opportunity to identify patients before symptoms appear. This is crucial for early intervention, delaying disease progression, and improving treatment outcomes. Some studies have shown that combining genome prediction models with clinical symptoms can more accurately predict individuals' risk of developing the disease (Oriol et al., 2019), providing a new direction for early intervention and treatment.

Genome prediction also provides a new way to gain a deeper understanding of the mechanisms of Alzheimer's disease. In the process of studying the genes that play a key role in predictive models, researchers can explore the biological functions of these genes in the development and progression of the disease, providing clues for the discovery of new therapeutic targets. By comprehensively understanding the genetic variations associated with Alzheimer's disease, a better understanding of the pathogenesis of the disease can be achieved, providing support for the development of precision medicine.

In summary, genome prediction has injected new vitality into Alzheimer's disease research and opened the door to the realization of individualized medicine. However, despite significant progress, a series of challenges still need to be faced, such as model complexity and data privacy issues. Future research needs to continuously improve genome prediction models, combining multi-source data to further improve prediction accuracy and reliability, and provide more powerful support for early intervention and treatment of Alzheimer's disease.

2.3 Case study analysis

In the field of Alzheimer's disease, large-scale genomics studies provide opportunities for in-depth understanding of the genetic basis of the disease.

Case study 1:
Leonenko et al. (2019) developed a genome prediction model using GWAS data, focusing on Alzheimer's disease. They successfully integrated a large amount of genetic information, enabling more accurate prediction of individuals' risk of developing Alzheimer's disease. Notably, by combining genome prediction with clinical information, researchers not only improved prediction accuracy, but also provided new ways to distinguish high-risk individuals and develop early intervention plans.

Case study 2:
The meta-analysis study conducted by Jansen et al. (2019) has broadened our understanding of the genetic basis of Alzheimer's disease. They identified a series of new genetic variants associated with the risk of the disease, involving multiple functional pathways. This study not only provides new markers for the optimization of genome prediction models, but also deepens our understanding of the mechanism of Alzheimer's disease.

Case study 3:
Tan et al. (2017) introduced a new polygenic hazard score (PHS) method that is associated with amyloid and tau protein deposition in Alzheimer's disease. By focusing on these biological markers, researchers not only improved the precision of genome prediction, but also provided a new perspective for understanding the biological mechanism of Alzheimer's disease.

The above studies, by integrating diverse genetic information and delving into biological markers, not only provide more comprehensive tools for individualized risk assessment, but also provide useful experiences for the development of future genome prediction models. These achievements lay a solid foundation for the prevention and treatment of Alzheimer's disease.

3 Genome Prediction and Its Association with Alzheimer's Disease

3.1 Comparison of research methods

In the research on the association between genome prediction and Alzheimer's disease, different research teams have adopted different methods to reveal the relationship between genetic factors and the disease. There are certain similarities and differences among these research methods, which are mainly reflected in the following aspects:

1) Selection and weight assignment of genetic markers

Different studies vary in the selection of genetic markers. Some studies focus on analyzing specific genes or gene regions, while others prefer to conduct comprehensive assessments through whole-genome approaches. Additionally, the weight assignment for different genes also varies among studies, which means that some studies may place more emphasis on the contribution of specific genes, while others may consider the role of multiple genes comprehensively.

2) Consideration of environmental factors

The degree of consideration of environmental factors also varies among studies. Some studies incorporate environmental factors into the model, attempting to comprehensively consider the interaction between genetics and the environment, while others may primarily focus on genetic factors and treat them as the core of the research.

3) Selection of datasets and sample size

The choice of datasets and sample size used in studies is also an important difference in research methods. Some studies may use data from different regions or ethnic groups to increase the external validity of the research, while others may focus on in-depth research within specific populations.

4) Analyzing the complexity of the model

The complexity of the analytical models used in the study also varies. Some studies use relatively simple statistical models, while others may use more complex machine learning algorithms to better capture the potential patterns in the genomic data.

Comparing the similarities and differences of research methods helps us to fully understand the diversity of genomic prediction research and provide guidance for future studies. By integrating different methods, we can build more comprehensive and robust genomic prediction models to better understand the genetic basis of Alzheimer's disease.

3.2 The application of the latest genome prediction technology in Alzheimer's disease research

With the continuous progress of technology, the latest genome prediction technology has ushered in a new chapter in Alzheimer's disease research. One remarkable technology is single-cell RNA sequencing (scRNA-seq). Through this technology, researchers canin-depth exploration the differences in gene expression in brain tissue at the single-cell level, Revealed the unique contribution of different cell types in the development of Alzheimer's disease. This fine resolution allows us to more comprehensively and accurately understand the pathological process of Alzheimer's disease.

In addition, the application of artificial intelligence (AI) has also brought revolutionary changes to genome prediction. Through the deep learning and pattern recognition capabilities of AI algorithms, researchers can discover hidden associations and patterns in large genomic datasets. In Alzheimer's disease research, AI can not only assist in analyzing the complex relationship between genes and diseases, but also predict the development trajectory of patient prognosis, providing higher-level guidance for precision medicine practice.

The application of these emerging technologies allows researchers to explore the information of genomics at a deeper and more comprehensive level in Alzheimer's disease research. The introduction of these advanced technologies not only expands our understanding of the mechanism of the disease, but also points to the future development of genome prediction. In this promising field, it ispromising to break through the limitations of

Alzheimer's disease research and provide more precise tools for the development of early diagnosis and treatment strategies.

3.3 Potential breakthroughs and discoveries

By deeply exploring the genome data, researchers may identify specific genetic variations that emerge in the early stages of Alzheimer's disease development. These early predictive markers have the potential to become tools for earlier diagnosis of Alzheimer's disease, providing a window for early intervention and treatment.

With a deeper understanding of the unique characteristics of individual genomes, doctors can develop more targeted treatment plans for patients. This means that in the future, it may be possible to develop more effective personalized medications and refined treatment plans that maximize therapeutic effectiveness and minimize the occurrence of side effects.

Additionally, with the widespread application of artificial intelligence in genomics, we are on the cusp of more intelligent genome prediction models. Through further optimization of machine learning algorithms, these models will be able to more accurately identify potential genetic associations and predict individual disease risks. This will provide clinicians with more reliable tools to better assist in decision-making and provide personalized medical advice.

These findings will not only deepen our understanding of the mechanisms underlying Alzheimer's disease, but also hold promise for future treatment and prevention strategies. By combining the latest technologies and interdisciplinary research methods, significant achievements are expected in the near future.

4 Summary and Outlook

Although genome prediction has made significant progress in Alzheimer's disease research, there are still a series of methodological challenges. The construction and optimization of models rely on large-scale genomic data. For complex diseases like Alzheimer's disease, larger and more diverse datasets are needed to improve the stability and generalization ability of models. Additionally, there are differences in methods and parameters used in different studies, which limits the consistency and comparability of results.

Alzheimer's disease is a complex disease involving multiple factors and genetic markers, so the interpretability of genome prediction models becomes another challenge. Even if models perform well, we still have limited knowledge about the specific roles of specific genetic markers in the disease mechanism. This leads to difficulties in explaining prediction results, limiting the feasibility of genome prediction in clinical applications.

Moreover, genome prediction models are typically based on population-level data, and there are significant individual differences. Personalized genome prediction models need to take into account individual lifestyles, environmental exposures, and other factors, which increases the complexity of data and interpretation.

With the deepening of genomics research, ethical and privacy issues become increasingly important. Genome prediction involves a large amount of highly sensitive genetic information, so privacy protection needs to be strengthened during data collection, storage, and sharing. At the same time, how to explain and present the results of genome prediction to individuals, as well as how to apply this information in clinical practice, also requires clearer ethical guidance.

It is also important to consider the social impact of genetic information. When conducting genome prediction, it may reveal information related to other diseases, traits, or family history, which may have potential impacts on individuals' employment, insurance, and other aspects. Establishing a sound ethical framework that safeguards individual rights while promoting scientific research is one of the important challenges facing current genome prediction research.

Despite the challenges faced by genome prediction in Alzheimer's disease research, there are still many promising future research directions. Personalized genome prediction models will be the focus of future research.

Researchers can also further explore the potential value of genome prediction in the early diagnosis and prevention of Alzheimer's disease. Additionally, with the continuous advancement of technology, the widespread application of new technologies such as whole-genome sequencing and single-cell sequencing will provide more abundant data for genome prediction research. The development of these technologies is expected to help researchers more comprehensively and deeply understand the genetic basis of Alzheimer's disease.

References

Ando K., Nagaraj S., Küçükali F., De Fisenne M.A., Kosa A.C., Doeraene E., and Leroy K., 2022, PICALM and Alzheimer’s disease: an update and perspectives, Cells, 11(24): 3994.
https://doi.org/10.3390/cells11243994

Bellenguez C., Küçükali F., Jansen I.E., Kleineidam L., Moreno-Grau S., Amin N., and Goldhardt O., 2022, New insights into the genetic etiology of Alzheimer’s disease and related dementias, Nature genetics, 54(4): 412-436.

Clark K., Leung Y.Y., Lee W.P., Voight B., and Wang L.S., 2022, Polygenic risk scores in Alzheimer’s disease genetics: methodology, applications, inclusion, and diversity, Journal of Alzheimer's Disease, 89(1):1-12.
https://doi.org/10.3233/JAD-220025

Fagundes S.D., Silva M.T., Thees M.F.R.S., and Pereira M.G., 2011, Prevalence of dementia among elderly Brazilians: a systematic review, Sao Paulo Medical Journal, 129: 46-50.
https://doi.org/10.1590/S1516-31802011000100009

Gao P., Ye L., Cheng H., and Li H., 2021, The mechanistic role of bridging integrator 1 (BIN1) in Alzheimer’s disease, Cellular and Molecular Neurobiology, 41(7): 1431-1440.
https://doi.org/10.1007/s10571-020-00926-y

Hu T., Chitnis N., Monos D., and Dinh A., 2021, Next-generation sequencing technologies: An overview, Human Immunology, 82(11): 801-811.
https://doi.org/10.1016/j.humimm.2021.02.012

Jansen I.E., Savage J.E., Watanabe K., Bryois J., Williams D.M., Steinberg S., and Posthuma D., 2019, Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk, Nature genetics, 51(3): 404-413.

Knopman D.S., Amieva H., Petersen R.C., Chételat G., Holtzman D.M., Hyman B.T., and Jones D.T., 2021, Alzheimer disease, Nature reviews Disease primers, 7(1): 33.
https://doi.org/10.1038/s41572-021-00269-y

Lambert S.A., Abraham G., and Inouye M., 2019, Towards clinical utility of polygenic risk scores, Human molecular genetics, 28(R2): R133-R142.
https://doi.org/10.1093/hmg/ddz187

Leonenko G., Sims R., Shoai M., Frizzati A., Bossù P., Spalletta G., and Escott-Price V., 2019, Polygenic risk and hazard scores for Alzheimer's disease prediction, Annals of clinical and translational neurology, 6(3): 456-465.
https://doi.org/10.1002/acn3.716

McGuire A.L., Gabriel S., Tishkoff S.A., Wonkam A., Chakravarti A., Furlong E.E., and Kim J.S., 2020, The road ahead in genetics and genomics, Nature Reviews Genetics, 21(10): 581-596.
https://doi.org/10.1038/s41576-020-0272-6

Nichols E., Steinmetz J.D., Vollset S.E., Fukutaki K., Chalek J., Abd-Allah F., .and Liu X., 2022, Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019, The Lancet Public Health, 7(2): e105-e125.
https://doi.org/10.1002/alz.051496

Oriol J.D.V., Vallejo E.E., Estrada K., Peña J.G.T., and Alzheimer’s Disease Neuroimaging Initiative, 2019, Benchmarking machine learning models for late-onset alzheimer’s disease prediction from genomic data, BMC bioinformatics, 20.
https://doi.org/10.1186/s12859-019-3158-x

Santos C.D.S.D., Bessa T.A.D., and Xavier A.J., 2020, Factors associated with dementia in elderly, Ciencia & saude coletiva, 25: 603-611.
https://doi.org/10.1590/1413-81232020252.02042018

Simonetti A., Pais C., Jones M., Cipriani M.C., Janiri D., Monti L., and Sani G., 2020, Neuropsychiatric symptoms in elderly with dementia during COVID-19 pandemic: definition, treatment, and future directions, Frontiers in psychiatry, 11: 579842.
https://doi.org/10.3389/fpsyt.2020.579842

Tan C.H., Hyman B.T., Tan J.J., Hess C.P., Dillon W.P., Schellenberg G.D., and Desikan R.S., 2017, Polygenic hazard scores in preclinical Alzheimer disease, Annals of neurology, 82(3): 484-488.
https://doi.org/10.1002/ana.25029

Wu J.W., Yaqub A., Ma Y., Koudstaal W., Hofman A., Ikram M.A., and Goudsmit J., 2021, Biological age in healthy elderly predicts aging-related diseases including dementia, Scientific reports, 11(1): 15929.
https://doi.org/10.1038/s41598-021-95425-5