



OPEN ACCESS

EDITED BY
Vivek Swarup,
University of California,
Irvine, United States

REVIEWED BY
Wei Xu,
Qingdao University Medical College,
China
Steven Estus,
University of Kentucky,
United States

*CORRESPONDENCE
Wan-Ping Lee
✉ Wan-Ping.Lee@Pennmedicine.upenn.edu

SPECIALTY SECTION
This article was submitted to
Alzheimer's Disease and Related Dementias,
a section of the journal
Frontiers in Aging Neuroscience

RECEIVED 19 October 2022
ACCEPTED 31 December 2022
PUBLISHED 08 February 2023

CITATION
Wang H, Wang L-S, Schellenberg G and Lee
W-P (2023) The role of structural variations in
Alzheimer's disease and other
neurodegenerative diseases.
Front. Aging Neurosci. 14:1073905.
doi: 10.3389/fnagi.2022.1073905

COPYRIGHT
© 2023 Wang, Wang, Schellenberg and Lee.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

The role of structural variations in Alzheimer's disease and other neurodegenerative diseases

Hui Wang^{1,2}, Li-San Wang^{1,2}, Gerard Schellenberg^{1,2} and Wan-Ping Lee^{1,2*}

¹Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, ²Penn Neurodegeneration Genomics Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Dozens of single nucleotide polymorphisms (SNPs) related to Alzheimer's disease (AD) have been discovered by large scale genome-wide association studies (GWASs). However, only a small portion of the genetic component of AD can be explained by SNPs observed from GWAS. Structural variation (SV) can be a major contributor to the missing heritability of AD; while SV in AD remains largely unexplored as the accurate detection of SVs from the widely used array-based and short-read technology are still far from perfect. Here, we briefly summarized the strengths and weaknesses of available SV detection methods. We reviewed the current landscape of SV analysis in AD and SVs that have been found associated with AD. Particularly, the importance of currently less explored SVs, including insertions, inversions, short tandem repeats, and transposable elements in neurodegenerative diseases were highlighted.

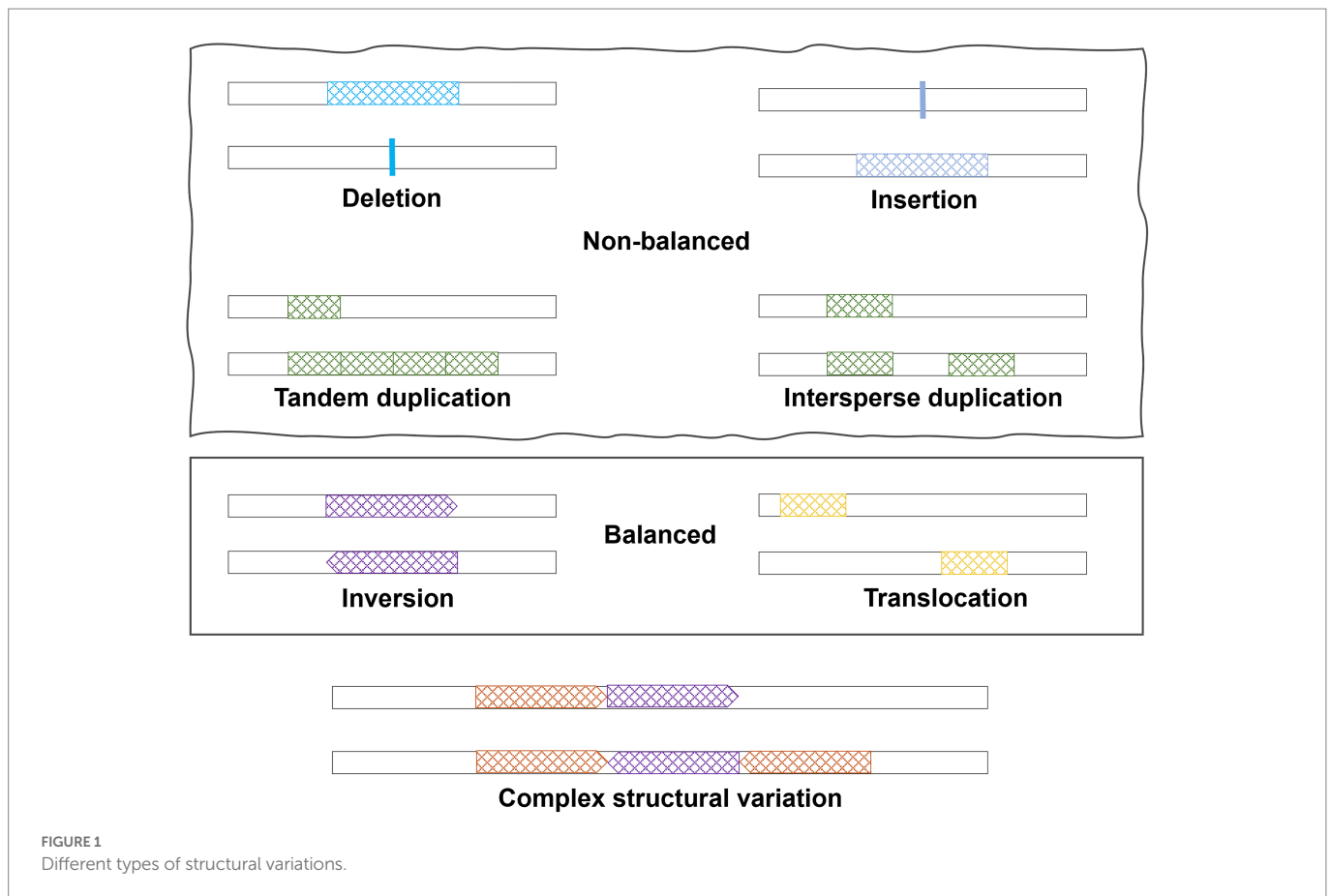
KEYWORDS

structural variations, Alzheimer's disease, next-generation sequencing, short-tandem repeats, transposable elements, copy number variations

Introduction

What is the genetic cause of Alzheimer's disease (AD)? The answer to this question has not changed much for the past decade. 10–20% of early-onset familial forms of AD are caused by mutations in *APP*, *PSEN1*, and *PSEN2* (Tanzi et al., 1987; Levy-Lahad et al., 1995; Sherrington et al., 1995). Genome-wide association studies (GWASs) have confirmed the role of *APOE* alleles in late-onset AD (LOAD) and identified dozens of other variants with small effects. A most recent GWAS revealed 42 new risk loci associated with AD with the odd ratio between 0.89 and 1.47 (Bellenguez et al., 2022). Common single nucleotide polymorphisms (SNPs) altogether were estimated to explain about 30% of phenotypic variance for AD (Lee et al., 2013; Ridge et al., 2013), which were calculated to have a heritability about 70% (Gatz et al., 2006). Epistasis and structural variants (SVs) were expected to account for the missing heredity in AD (Bertram et al., 2010; Raghavan and Tosto, 2017). However, there was only a limited increase in predicting the status of AD when epistasis was incorporated into polygenic risk (Wang et al., 2020, 2021). Compared to SNPs, SVs are estimated to contribute equally or more to genetic variations by total nucleotide content (Feuk et al., 2006). From the previous study, SVs account for 17.2% of strongly deleterious rare variants in the human genome (Abel et al., 2020). Though SVs were less studied in the etiology of diseases due to challenges in characterizing SVs accurately and exhaustively from array and short-read data, they are likely to be an important component of the missing heritability in AD.

SVs are large genomic variations (>50 bp), including copy number variations (CNVs, i.e., deletions and duplications), insertions, inversions, translocations, and complex combinations



(Figure 1). Deletions, duplications, and insertions are called non-balanced SVs as they increase or decrease the dosage of specific genes, while inversions and translocations are balanced SVs. Compared to SNPs, SVs may have a larger impact on the human genome and can alter not only protein coding sequences but also the dosage of a specific exon or entire gene. Considering the numerous publications of GWAS for SNPs, SVs in human diseases have been understudied due to limitations on SV detection methods. In this review, we started with a brief introduction of technologies and methods, which are commonly used to detect SVs. Subsequently, a summary of the current understandings of SVs in AD pathogenesis was provided. The review focused on the important roles of specific types of SVs in AD and neurodegenerative diseases and how recent sequencing technologies may help bring new insights.

Genome-wide detection of structural variants

Array-based methods

Large SVs at cytogenetic level (~2–3 Mb or above) can be identified by karyotyping, such as chromosomal aneuploidies in Down's Syndrome (Jacobs et al., 1959). Later, fluorescence *in situ* hybridization (FISH) (Pinkel et al., 1986) allows detection of smaller deletions and duplications as well as translocations using fluorescent probes that only bind to specific genes or chromosome regions. Based on FISH, comparative genomic hybridization (CGH; Kallioniemi

et al., 1992) was developed to produce a map of DNA sequence copy number throughout the whole genome. A broad application of CGH as a diagnosis tool requires higher resolution and simpler procedures, which leads to development of microarray-based comparative genomic hybridization (aCGH; Solinas-Toldo et al., 1997; Vissers et al., 2003; Bejjani and Shaffer, 2006) for a finer analysis of genomic CNVs. Since 2010, aCGH has been the first-tier cytogenetic test in place of G-banded karyotyping for patients with unexplained developmental delay or intellectual disability, autism spectrum disorders, and congenital anomalies (Manning et al., 2010; Miller et al., 2010). In aCGH, array can be built with DNA sequences whether from oligonucleotides, cDNAs, or bacterial artificial chromosomes. Then, reference and test DNA labeled with two different fluorophores probes are hybridized to the DNA sequences on the array (Bejjani and Shaffer, 2006). Comparing the fluorescence intensity between case and control, genomic gains or losses can be attained simultaneously. The disadvantage of aCGH is that it cannot provide the absolute copy number of a specific gene or region.

As the wide application of GWAS, numerous algorithms, such as PennCNV (Wang et al., 2007) and QuantiSNP (Colella et al., 2007), were developed to infer CNVs directly from SNP array data. Winchester et al. (2009) provided a comprehensive summary of existing algorithms for CNV inferred from SNP array. CNV inferred by SNP array are not as accurate as aCGH and usually cannot be used to detect small variants (<30 Kb) (Zhang et al., 2014) as the density of SNP probes is sparse for most genomic regions. However, the advantage is that CNV analysis can be performed directly on existing array data that were originally intended for GWAS. In addition to CNVs, SNP-array analysis is also

possible to infer certain balanced SV, such as inversions from linkage disequilibrium pattern (Cáceres et al., 2012).

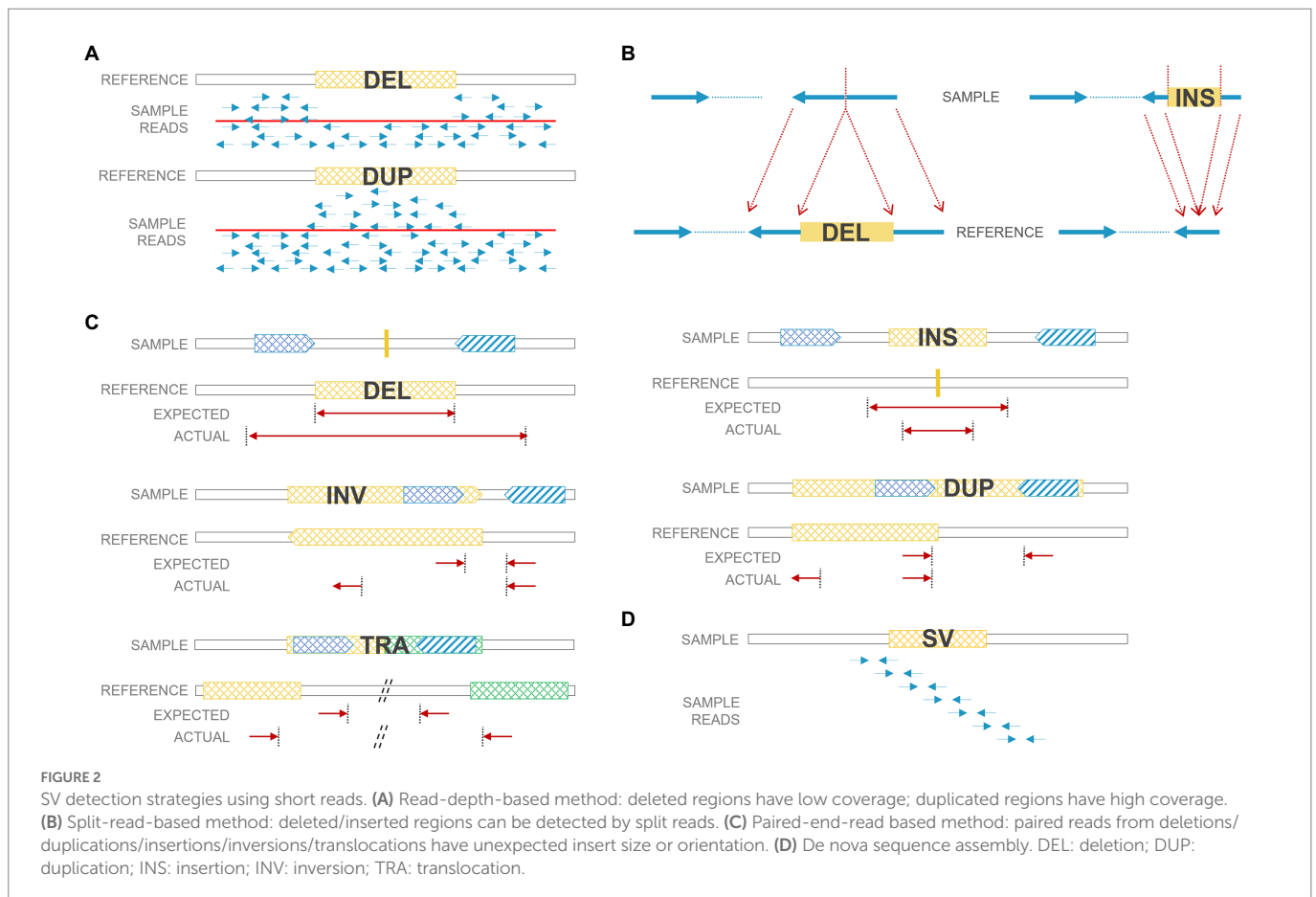
Short-read sequencing

As next-generation sequencing (NGS), especially paired-end sequencing, has become widely affordable and available, researchers have been able to detect SVs in a finer resolution and explore balanced SVs, including insertions and inversions that were often ignored before. However, the accuracy of calling SVs and the precision of breakpoints are still inadequate due to certain limitations of the short-read technology. There were numerous algorithms developed for SV detection, and a review of SV calling algorithms for NGS was summarized previously (Lin et al., 2015; Guan and Sung, 2016). Technically, there are four main strategies for SV calling: read depth, paired-end reads, split reads, and *de novo* assembly (Figure 2). 1. Read-depth-based algorithms mainly focus on CNV identification through comparison of the read depth of a specific region and the read depth of its surroundings (or average depth); 2. Paired-end-read-based algorithms can identify CNVs as well as insertions, inversions, and translocations. It detects abnormal insert size, alignment orientation (for duplications and inversions), or alignment locations (for translocations) between the two ends of a paired-end read; 3. Split-read-based methods detect SVs from reads that cross SV breakpoints. Along with breakpoints, reads can be split and mapped against the reference genome encompassing SVs. Therefore, theoretically, split-read methods can provide single-nucleotide resolution for SV breakpoints. However,

breakpoints regularly harbor short repeats that bring challenges when performing split-read strategy for SV detection. Also, higher coverage is needed to accumulate enough split reads for SV calling; 4. *De novo* sequence assembly reassembles contigs and then compares contigs to references. It is the most accurate and not biased by reference sequences, but it needs high coverage (or prone to assembly errors) and induces high computational cost. To increase the sensitivity in SV calling, several strategies were usually combined together to form a hybrid algorithm, such as LUMPY (Layer et al., 2014) and Manta (Chen et al., 2016).

Long-read sequencing and other technologies

Other than NGS, a few new methods and techniques can be applied to a more accurate SV detection, including single-molecule long-read sequencing by Pacific Biosciences and Oxford Nanopore Technologies, as well as new platforms such as Illumina Infinity. Particularly, long-read sequencing can help to solve regions that have few or low-mappable reads due to high GC content or other chemical issues, and also regions that have a low mapping quality due to repetitive sequences (Jain et al., 2018; Ebbert et al., 2019). Despite higher error rate on each nucleotide, long-read sequencing can provide higher mappability in those “problematic regions,” have ability to span entire SVs, and offer sufficient long split reads to lead a better tool for SV detection. Overall, long reads and phased genome generated by long-read sequencing enable more precise detection of breakpoints and more SVs compared to short-read sequencing (De Coster and Van Broeckhoven, 2019).



Furthermore, optical mapping detects SVs from *de novo* assemblies created by imaging intact single molecules of native-state DNA (Chan et al., 2018). Synthetic long-read techniques (including Illumina TruSeq) generate low-error local assembly of short-read data through specific library preparations (McCoy et al., 2014; Zheng et al., 2016) and can better resolve repetitive regions and phase haplotypes for SV detection. By tagging nascent strand using bromodeoxyuridine (thymidine analog) during replication, Strand-seq is able to sequence individual DNA strand through removing the nascent strand following photolysis induced by UV light (Falconer et al., 2012; Ho et al., 2020). The directionality obtained from single strand sequencing makes Strad-seq particularly well-suited for inversion discovery and haplotype phasing. Additionally, Hi-C can be used to infer interchromosomal translocations and large intrachromosomal SVs by abnormal chromatin interactions around the SV breakpoints (Lieberman-Aiden et al., 2009; Wang et al., 2022).

Copy number variants in Alzheimer's disease

Overall burden of CNVs in AD

In autism and schizophrenia, many studies have confirmed large (>100 Kb) rare (<1%) CNV (usually deletion) burden in patients versus controls (Walsh et al., 2008; Kirov et al., 2009a; Buizer-Voskamp et al., 2011; Kushima et al., 2018; Maury et al., 2022). However, there were mixed results in the evaluation of overall CNV burden in AD studies. Some earlier analysis based on SNP array found no significant difference in CNV rate and CNV size between AD cases and controls (Ghani et al., 2012; Swaminathan et al., 2012b). Guffanti et al. (2013) found overrepresentation of large heterozygous deletions in cases (p -value <0.0001) using 459 AD/mild cognitive impairment (MCI) cases and 181 health controls in ADNI. From CNVs called on NGS data, even though no significant difference in CNV count or length between AD cases and controls were found (Lee et al., 2021; Ming et al., 2021), non-Hispanic-white AD cases showed more duplications, and Hispanic AD cases showed larger deletions when burden analysis was stratified by ethnicity (Lee et al., 2021). Due to the fact that only large CNVs can be detected from SNP array and false positive CNVs detection is a major concern for short read data, future studies involving more robust and comprehensive CNV callings are needed to clarify the overall burden of CNVs in AD.

CNVs in early-onset AD

For early-onset familial AD, hundreds of pathogenic mutations in *APP*, *PSEN1*, and *PSEN2* have been discovered. In a study from France that included 12 unrelated individuals with autosomal dominant early-onset AD (EOAD), 70 unrelated individuals with familial LOAD, and 100 health controls, the *APP* duplications ranging from 0.58 to 6.37 Mb were found in five individuals with autosomal-dominant EOAD (Table 1; Rovelet-Lecrux et al., 2006). In other studies of Europeans, however, the *APP* duplication was identified as a relatively rare cause of autosomal dominant EOAD (Sleegers et al., 2006; Blom et al., 2008; Hooli et al., 2012). In samples with Japanese ancestry, two unrelated early-onset familial AD families were found to harbor the *APP* duplication (with SV size about 4.2 and 0.7 Mb) among 25 families with

familial AD and 11 sporadic EOAD cases (Kasuga et al., 2009). Overall, duplications in *APP* are rare, but indeed an important cause of autosomal dominant EOAD.

Other than the widely recognized *APP* duplications, deletions of exon 9 in *PSEN1* have been discovered in families affected by a variant of AD with spastic paraparesis and unusual plaques (Table 1; Crook et al., 1998; Smith et al., 2001), and 10 novel CNVs overlapping a set of genes, including *A2BP1*, *ABAT*, *CDH2*, *CRMP1*, *DMRT1*, *EPHA5*, *EPHA6*, *ERMP1*, *EVC*, *EVC2*, *FLJ35024*, and *VLDLR*, were found to co-segregate with disease status from 261 early-onset familial AD families (Hooli et al., 2014).

Array-based analysis

Using CNVs inferred from SNP array, Heinzen et al. (2010) reported a large 500 Kb duplication located at 15q13.3 in 6 of 276 cases and 1 of 322 controls (Table 1, $p=0.053$). The duplicated region covers entire *CHRNA7* gene and was confirmed to be associated with schizophrenia and epilepsy (Stefansson et al., 2008). Interestingly, Swaminathan et al. (2011) also identified a 30 Kb deletion from ADNI that overlaps with *CHRFAM7A*, which is *CHRNA7* (Exons 5–10) and *FAM7A* (Exons A–E) fusion, and appears in 4 of 471 AD/MCI cases but not in 184 health controls (Table 1). In the same study, CNVs in *CSMD1*, *HNRNPCL1*, and *SLC35F2* were case-specific with a nominal $p < 0.05$, though no significant signal was reported after adjusting for multiple tests (Table 1; Swaminathan et al., 2011). In 2012, Swaminathan et al. (2012b) extended the study and performed similar analysis on 882 unrelated non-Hispanic Caucasian participants in the National Institute of Aging LOAD/National Cell Repository for AD (NIALOAD/NCRAD) Family Study. Again, no genome-wide significant signal was observed. However, CNVs in five AD candidate genes (*ATXN1*, *HLA-DPB1*, *RELN*, *DOPEY2*, and *GSTT1*) only showed in cases in both NIALOAD/NCRAD and ADNI (Table 1). When TGen cohort consisting of 728 cases and 438 controls was used to validate the results from NIALOAD/NCRAD and ADNI, CNVs in AD candidate genes, including a number of previously reported regions (*CHRFAM7A*, *RELN*, and *DOPEY2*) as well as a new gene (*HLA-DRA*), were identified (Table 1; Swaminathan et al., 2012a). Using the four algorithms to detect CNVs from 1,103 samples, Ghani et al. (2012) identified a 470 Kb duplication in 15q11.2 (encompassing *TUBGCP5*, *CYFIP1*, *NIPA2*, *NIPA1*, and *WHAMML1*) appearing in 10 cases (2.6%) and three controls (0.8%) with nominal significance ($p=0.037$), and indicated that CNVs in *NRXN1*, *CNTNAP2*, *PTPRD*, *NDUFAF2*, and *CNTN6* were also worthy of further study (Table 1). Notably, for the deletion in *NRXN1*, several other reports suggested that it could increase the risk of developing Schizophrenia (Kirov et al., 2009b; Maury et al., 2022). Combining gene expression with dosage information from SNP array, an 8 Kb deletion containing a PAX6-binding site on the upstream of *CREB1* was associated with AD (Table 1; Li et al., 2012). In a study comparing 33 EOAD cases with 212 LOAD cases and 1,078 controls using high-resolution aCGH, seven singleton CNVs were reported from EOAD samples (Table 1; Rovelet-Lecrux et al., 2012).

None of genes mentioned above reached genome-wide significance after adjusting for multiple tests. Two studies found genome-wide significant CNVs using alternative phenotypes (Age of onset and AD with psychosis): Szigeti et al. (2013) identified five significant regions

TABLE 1 Copy number variants associated with Alzheimer's disease.

Gene	Location	Size	Consequence	Method	CNV	Sample	Evidence	References
APP	21q21.3	0.57–6.37 Mb	Gene duplication	PCR	Duplication	ADEOAD: 12, familial LOAD: 70, control: 100	CNV in 5 ADEOAD families	Rovelet-Lecrux et al. (2006)
PSEN1	14q24.2	-	Deletion of exon 9	PCR	Deletion	1 family: 60 individuals	CNV in 17 of 60	Crook et al. (1998)
CHRNA7	15q13.3	500 Kb	Gene duplication	SNP Array	Duplication	AD: 331, control: 368	$p=0.053$ (six cases, one control)	Heinzen et al. (2010)
CHRFAM7A	15q13.2	30 Kb	Deletion of multiple exons	SNP Array	Deletion	AD: 222, MCI: 136, control: 143	Case only	Swaminathan et al. (2011)
CYFIP1	15q11.2	470 Kb	Gene duplication	SNP Array	Duplication	AD: 559, control: 554	$p=0.037$ (10 cases, three controls)	Ghani et al. (2012)
CR1	1q32.2	18 Kb	Additional C3b/C4b binding sites	PCR	Duplication	AD: 1039, control: 844	$p<0.05$	Brouwers et al. (2012)
SULT1A3/4	16p11.2	-	Gene duplication	PCR	Duplication	AD: 91, control: 172	$p=0.011$	Butcher et al. (2018)
AMY1A	1p21.1	-	Gene duplication	PCR	Duplication	AD: 247, control: 5175	$p<0.05$	Byman et al. (2020)
CPNE4	3q22.1	3.6 Kb	Intragenic deletion	SNP Array	Deletion	AD: 375, control: 192	FDR = 0.033	Szigeti et al. (2013)
APC2	19p13.3	-	-	SNP Array	Duplication	AD (psychosis): 496, AD (intermediate psychosis): 639, AD: 156	7.20×10^{-10}	Zheng et al. (2015)
CREB1	2q33.3	8 Kb	Upstream of CREB1 (Pax-6 binding site)	Expression/ SNP Array	Deletion	AD: 22, control: 15 (DEG); AD: 1230, control: 936 (CNV)	p (DEG) < 0.001; p (CNV) = 0.008	Li et al. (2012)
A2BP1, ABAT, CDH2, CRMP1, DMRT1, EPHA5, EPHA6, ERMP1, EVC, EVC2, FLJ35024, VLDLR				SNP Array	CNVs	261 families: 1,009 individuals	Case only	Hooli et al. (2014)
CSMD1, HNRNPCL1, SLC35F2				SNP Array	CNVs	AD: 222, MCI: 136, control: 143	$p<0.05$	Swaminathan et al. (2011)
ATXN1, HLA-DPB1, RELN, DOPEY2, GSTT1				SNP Array	CNVs	AD: 711, control: 171	Case only	Swaminathan et al. (2012b)
CHRFAM7A, RELN, DOPEY2, HLA-DRA				SNP Array	CNVs	AD: 728, control: 438	$p<0.05$	Swaminathan et al. (2012a)
MAG11, NLRP8, KLK6, MEOX2, SLC30A3, FPR2, SVOP				aCGH	CNVs	ADEOAD: 23, sporadic EOAD: 12, other AD: 912, control: 1078	EOAD only singletons	Rovelet-Lecrux et al. (2012)

AD, Alzheimer's Disease; ADEOAD, Autosomal Dominant Early-Onset Alzheimer's Disease; CNVs: Copy Number Variation; DEG, Differentially Expressed Gene; EOAD, Early-Onset Alzheimer's Disease; LOAD, Late-Onset Alzheimer's Disease; MCI, Mild Cognitive Impairment.

(FDR < 0.05) ranging from 3.6 to 24.8 Kb that were related to the age of onset in AD, including an intragenic deletion in *CPNE4* (Table 1). Using 496 AD cases with psychosis, 639 AD cases with intermediate psychosis

and 156 AD cases without psychosis, Zheng et al. (2015) found a duplication (odds ratio = 0.42; $p = 7.2 \times 10^{-10}$) in the *APC2* that is protective against developing psychosis in AD (Table 1).

Whole-genome-sequencing-based analysis

Just like studies done on SNP arrays, association analysis of SVs detected from NGS yielded no genome-wide significant signal. There could be a few explanations for that. First, the sample size of studies using NGS is usually smaller than SNP arrays, although more SVs are expected from using NGS. Second, even joint genotyping can increase sample sizes, joint genotyping is tough to be properly done for SVs due to lack of well-aligned breakpoints. Third, repetitive genomic regions cause artificial alignments that would lead to false positive SV detection and introduce noises of association analysis.

While association between SVs and AD status reveals no significant signal, using 1,411 samples in MSBB and ROSMAP, Ming et al. (2021) found that the AD-specific CNVs showed distinct functional annotations compared to MCI-specific and normal-specific CNVs, such as glucuronosyltransferase activity, cellular glucuronidation, and neuron projection. Moreover, Vialle et al. (2022) performed an SV-xQTL analysis and identified more than 3,200 SVs that were correlated with histone modifications, gene expression, splicing, or protein abundance in postmortem brain tissues, providing a valuable resource for functional study of SVs.

Single-cell whole genome sequencing (scWGS) has been applied to study the aneuploidy in AD patients (van den Bos et al., 2016). Since neurons are post-mitotic, earlier studies of increased overall aneuploidy in AD patients were based on the analysis of metaphase cells in whole peripheral blood (Ward et al., 1979). The application interphase FISH allowed detection of aneuploidy in neuronal cells in human brains. However, interphase FISH can only target a limited number of chromosomes in one cell and is intrinsic noisy, and thus studies often showed different results. Iourov et al. (2009) reported chromosome 21-specific aneuploidies in the cerebral cortex of AD patients, while other studies indicated aneuploidies caused by increased tetraploid neurons in AD patients due to a full S phase without initiation of mitosis (Yang et al., 2001; Mosch et al., 2007). With the advent of scWGS, all chromosomes in a cell can be analyzed, and each chromosome in a cell was probed thousands of times by different sequencing reads. Moreover, unlike interphase FISH, analysis of aneuploidy by scWGS would not be affected by tissue sectioning or other possible causes of artifacts. van den Bos et al. (2016) used scWGS for studying aneuploidies of neuronal cells and found no significant difference between AD cases and controls. Besides, giving the wide application of scWGS in human genetics and disease etiology (Zhang et al., 2019; Porubsky et al., 2021; Salehi et al., 2021), the discovery of somatic SVs in single-cell resolution in brains of AD patients would definitely yield valuable new insights about AD pathogenesis.

Targeted analysis of CNVs on specific genes in AD

Complement receptor gene 1 (*CR1*) was identified and replicated as a causal gene for AD by several GWASs (Harold et al., 2009; Lambert et al., 2009). However, the specific mechanism of *CR1* disruption in AD remains unclear. The significant SNP (rs4844610) was in an LD block which spans nearly the entire gene *CR1*, except for the first and last exons. Within the LD block, there is an 18 Kb low-copy repeat (LCR) which represents the same signal as the SNP by conditional regression analysis (Table 1; Brouwers et al., 2012). Generally, LCRs usually lead to genome instability due to non-allelic homologous recombination

(Erdogan et al., 2006). In *CR1* protein, the LCR can affect protein function by creating additional binding sites for complement components C3b and C4b (Brouwers et al., 2012). Therefore, the LCR CNV is likely to explain the significant association within this LD block.

SULT1A3/4 are important enzymes in the metabolism of catecholamines which are involved in many neurodegenerative diseases. Butcher et al. (2018) evaluated association of copy number of *SULT1A3/4* and the risk of AD and Parkinson's Disease (PD; Table 1). For those individuals with less than four copies of *SULT1A3/4*, their ages of onset for AD were earlier, and they were more likely to develop AD (odds ratio = 1.69), but the association with PD was not significant.

Byman et al. (2018) found evidences supporting the role of α -amylase involvement in AD pathology. Subsequently, they further studied relationship between *AMY1A* copy number with AD status and memory performance (Byman et al., 2020). There is no significant difference in *AMY1A* copy number between cases and controls. However, individuals with high copy number of *AMY1A* (≥ 10) showed significantly lower hazard ratio compared to reference (Table 1). The full list of CNVs associated with AD is listed in Table 1.

Other structural variants in neurodegenerative diseases

Inversions, insertions, and complex SVs

Compared to CNVs, insertions, inversions, and complex SVs are less studied due to the limitation of array-based analysis, although they are important causal factors in the pathogenesis of neurodegenerative diseases. Sato et al. (2009) identified a 2.5–3.8 Kb insertion containing a long (TGGAA)_n stretch in patients with spinocerebellar ataxia type 31 (Table 2); the length of the insertion was inversely correlated with age onset of the disease. Prion diseases are often caused by a conversion of normal prion protein (PrP^C) into a proteinase K resistant form of PrP (PrP^{Sc}). Insertion of 2–12 octapeptide repeats between codon 51–91 of *PRNP* can cause genetic form of prion disease (Table 2; Rossi et al., 2000; Moore et al., 2001; Kim et al., 2018). These octapeptide insertion can cause rapid binding between PrP molecules, therefore increasing the rate of PrP^{Sc} formation (Moore et al., 2006). In a study of SVs in amyotrophic lateral sclerosis (ALS) genes, an inversion in the *VCP* gene and an insertion in *ERBB4* gene were identified to be able to increased risk of ALS besides the repeat expansion in *C9ORF72* from whole genome sequencing of 6,500 individuals (Table 2; Al Khleifat et al., 2022).

Non-canonical or complex SVs are usually hard to decipher as they are typically composed of multiple breakpoint junctions and cannot be characterized as a single canonical SV type, but they may harbor important disease-causing variants. For example, the region of 17q21.31 containing *MAPT* has two major haplotypes: H1 and H2. The H2 haplotype is characterized by a ~900 Kb inversion flanked by two duplication blocks, and tagged by a 238 bp deletion between exons 9 and 10 of *MAPT* and a few SNPs inside (Baker et al., 1999). The H2 haplotype is positively selected in Europeans and is completely absent or extremely rare in other ethnic groups (Pittman et al., 2006). Nearly all individuals carrying the H2 haplotype were in the same structural form (H2.α2.γ2), indicating it may derive from a single founder (Boettger et al., 2012). In contrast, the H1 haplotype is more variable and exists in all ethnic groups. As for the origin of H1 and H2, one study reported that they do not follow a precursor-product relationship and cannot be derived

TABLE 2 Other structural variants in neurodegenerative diseases.

Type	SVs	Location	Gene	Risk allele	Inheritance	Disease
STR	(CAG) _n	4p16.3	HTT	>39	Autosomal dominant	Huntington's disease
STR	(CTG) _n	16q24.2	JPH3	>39	Autosomal dominant	Huntington's disease-like 2
STR	(CGG) _n	Xq27.3	FMR1	>200	X-linked dominant	Fragile X syndrome
STR	(GAG) _n	12p13.31	ATN1	>48	Autosomal dominant	Dentatorubral–pallidolusian atrophy
STR	(CAG) _n	6p22.3	ATXN1	>39	Autosomal dominant	Spinocerebellar ataxia type 1
STR	(CAG) _n	12q24.12	ATXN2	>36	Autosomal dominant	Spinocerebellar ataxia type 2
STR	(CAG) _n	14q32.12	ATXN3	>60	Autosomal dominant	Spinocerebellar ataxia type 3
STR	(CAG) _n	19p13.2	CACNA1A	>18	Autosomal dominant	Spinocerebellar ataxia 6
STR	(CAG) _n	3p14.1	ATXN7	>36	Autosomal dominant	Spinocerebellar ataxia 7
STR	(CTG) _n	13q21.33	ATXN8OS	>53	Autosomal dominant	Spinocerebellar ataxia 8
STR	(ATTCT) _n	22q13.31	ATXN10	>800	Autosomal dominant	Spinocerebellar ataxia 10
STR	(CAG) _n	5q32	PPP2R2B	>54	Autosomal dominant	Spinocerebellar ataxia 12
STR	(CAG/CAA) _n	6q27	TBP	>40	Autosomal dominant	Spinocerebellar ataxia 17
STR	(GGCCTG) _n	20p13	NOP56	>650	Autosomal dominant	Spinocerebellar ataxia 36
STR	(GAA) _n	9q21.11	FXN	>70	Autosomal recessive	Friedreich's ataxia
STR	(GGGGCC) _n	9p21.2	C9ORF72	>30	Mainly in familial ALS	Amyotrophic lateral sclerosis
STR	(CAG) _n	12q24.12	ATXN2	>26	Increasing risk	Amyotrophic lateral sclerosis
STR	(CA) _n	8q21.13	STMN2	24	Increasing risk	Amyotrophic lateral sclerosis
STR	(GGGGCC) _n	9p21.2	C9ORF72	>30	Mainly in familial FTD	Frontotemporal dementia
STR	(CAG) _n	12q24.12	ATXN2	>36	Autosomal dominant	Parkinson's disease
STR	(GGGGCC) _n	9p21.2	C9ORF72	>60	Rare cause	Parkinson's disease
STR	(AAGGG) _n	4p14	RFC1	>144	Rare cause	Parkinson's disease
STR	repeat expansion in ATXN3, ATXN10, and TBP				In PD patients	Parkinson's disease
STR	(CAG/CAA) _n	6q27	TBP	>38	Increasing risk	Multiple system atrophy
VNTR	25 bp motif	19p13.3	ABCA7	>228	Increasing risk	Alzheimer's disease
VNTR	14 bp motif	11p15.5	INS	Class III allele homozygotes	Earlier onset	Alzheimer disease
VNTR	69 bp motif	18q21.31	WDR7	–	Increasing risk	Amyotrophic lateral sclerosis
VNTR	21 bp motif	2q35	XRCC5	2R allele	Increasing risk	Multiple Sclerosis
VNTR	86 bp motif	2q13	IL1RN	IL1RN*2 allele	Earlier onset	Wilson's disease (neuropsychiatric form)
SVA	SVA	2q14.3	BIN1	–	Increasing risk	Alzheimer's disease
SVA	SVA	16p11.2	BCKDK	–	Increasing risk	Parkinson's disease
SVA	(CCCTCT) _n	Xq13.1	TAF1	>34	X-linked	X-linked dystonia-parkinsonism
SVA	SVA	17q21.31	MAPT	–	Increasing risk	Progressive supranuclear palsy
Alu	PolyT	19q13.32	TOMM40	18–30	Increasing risk	Alzheimer disease
Insertion	Octapeptide repeat insertion	20p13	PRNP	2–12 octapeptide repeats	Autosomal dominant (>4 repeats)	Prion disease
Insertion	(TGGA) _n	16q21	BEAN1	Repeat insertion	Autosomal dominant	Spinocerebellar ataxia 31
Insertion	Insertion	2q34	ERBB4	–	Increasing risk	Amyotrophic lateral sclerosis
Inversion	Inversion	9p13.3	VCP	–	Increasing risk	Amyotrophic lateral sclerosis
Complex SVs	~1 Mb haplotype	17q21.31	MAPT	H1	Increasing risk	Progressive supranuclear palsy

directly from each other (Hardy et al., 2005), while another study preferred a H2-like ancestor (Zody et al., 2008).

Many studies have reported the association of the inverted H2 haplotype with reduced risk of a range of neurodegenerative disease, including progressive supranuclear palsy (PSP) (Table 2; Baker et al., 1999), frontotemporal disorders (FTD; Baker et al., 1999), AD (Allen et al., 2014), and PD (Zabetian et al., 2007). Due to the important role of *MAPT* in tau pathology, variants in *MAPT* were among the most studied ones in this complex region. *MAPT* can produce six major tau isoforms by the alternative splicing of exons 2, 3, and 10. Alternative splicing of exon 10 can lead to imbalance of three microtubule-binding repeats (3R) tau (exclusion of exon 10) and four microtubule-binding repeats (4R) tau (inclusion of exon 10) ratio which were associated with several tauopathies in brain (Pittman et al., 2006). It was showed that H1 expresses more *MAPT* exon 10 mRNA compared to H2 (Caffrey et al., 2006), conforming to the increased 4R tau isoforms in PSP (Arai et al., 2001). The 238 bp deletion between exons 9 and 10 was considered to influence the alternative splicing of exon 10, therefore, affecting the risk of developing PSP (Baker et al., 1999). In addition, four highly correlated SVs [three deletions and one SINE-VNTR-Alu (SVA)] tagging H1/H2 haplotype were identified and may regulate nearby gene expression (Vialle et al., 2022).

Other than variants in LD with the H2/H1 haplotype, several deleterious rare SVs were also identified in this region. Microdeletions (~600 Kb) encompassing *MAPT* and *CRHR1* in 17q21.37 were identified in three individuals with intellectual disability (Koolen et al., 2006). A partial deletion of exons 6 to 9 of *MAPT* causing a truncated tau isoform was detected in a FTD patient (Rovelet-Lecrux et al., 2009). Duplications (~460 Kb) spanning *MAPT* were found in two PSP patients from a total of 283 PSP patients (Chen et al., 2019). A list of neurodegenerative diseases caused by insertions, inversions, and complex SVs were displayed in Table 2.

Short tandem repeats and variable number of tandem repeats

Short-tandem repeats (STRs) are repeating DNA sequences of 2–6 base pairs in length, which are often referred as microsatellites or simple sequence repeats (SSRs) when used in different circumstances. Many neuropathological diseases were caused by STRs, such as Huntington's disease, Fragile X syndrome, Dentatorubral-pallidolusian atrophy, and several spinocerebellar ataxias (Ryan, 2019). Huntington's disease is a progressive disorder that interrupts mood, movement, and intellectual abilities. It is an autosomal dominant disease caused by a trinucleotide CAG repeat expansion (from 36 to 120 CAG repeats in patients) in the *HTT* (Huntingtin) gene, which leads to a polyglutamine stretch causing dysfunction in the *HTT* protein (Table 2; Penney et al., 1997). In addition to CAG repeats in *HTT*, increased copy number of *SLC2A3* can increase the level of GLUT3, therefore, delaying the age onset of Huntington's disease. The experimental validation showed that increased dosage of *Drosophila melanogaster* homologue *Glut1* ameliorated Huntington's disease related phenotypes in fruit flies (Vittori et al., 2014). Another CTG repeat expansion in *JPH3* was causal for a similar autosomal dominantly inherited disease, i.e., Huntington disease-like 2 (Table 2; Stevanin et al., 2003).

Fragile X syndrome is a rare neurodegenerative disease that causes a range of developmental problems, including learning

disabilities, and cognitive impairment. Normally, people have less than 45 CGG repeats in the *FMR1* gene, while patients with Fragile X syndromes have more than 200 CGG repeats, causing failure in making FMRP protein (Table 2; Kremer et al., 1991). Dentatorubral-pallidolusian atrophy is a neurodegenerative disease characterized by variable combinations of myoclonus, epilepsy, ataxia, choreoathetosis, and dementia, and its clinical presentation correlates with the number of CAG repeats in *ATN1* (Table 2; Koide et al., 1994).

Ataxias were characterized by gait ataxia and other cerebellar signs due to progressive loss of nerve cells. Hereditary ataxias can be grouped into three categories: the autosomal dominant ataxias, also called spinocerebellar ataxias (SCA), autosomal recessive cerebellar ataxias (ARCA), and X-linked ataxias. SCA is the most common form of hereditary ataxias, and there are more than 30 genetic causes for SCA (Shakkottai and Fogel, 2013), many of which were caused by STRs in genes. For example, SCA1/SCA2/SCA3 are caused by an expanded CAG trinucleotide repeat in the *ATXN1/AXTN2/AXTN3* gene (Table 2; Banfi et al., 1994; Bürk et al., 1996; Damji et al., 1996; Stevanin et al., 1996). Simple repeat CAG/CAA (>38) in *TBP* is the cause of SCA17 (Nakamura et al., 2001) and is associated with increased risk of multiple system atrophy (Wernick et al., 2021). Friedreich's ataxia is one of the most common forms of ARCA. Patients with Friedreich's ataxia were caused by abnormal copy number of GAA trinucleotide repeat in *FXN* gene. Normally, there are less than 30 GAA repeats in *FXN*, while GAA repeats 70 to more than 1,000 times in patients with Friedreich's ataxia (Table 2; Campuzano et al., 1996; Anheim et al., 2012). The higher copy number of GAA was associated with more severe and faster evolving disease form (Bhidayasiri et al., 2005). Other than STRs, CNVs in *SACS*, *SYNE1*, *ADCK3*, and *SETX* were found to be potential causal genes for ARCA (Campuzano et al., 1996; Anheim et al., 2012; Cheng et al., 2021).

Currently, the association of expanded GGGGCC repeat in *C9ORF72* with FTD and ALS (DeJesus-Hernandez et al., 2011) has been widely recognized (Table 2). Besides mutations in *SOD1*, *TARDBP*, and *FUS*, GGGGCC repeat in *C9ORF72* is the main cause of ALS and accounts for 30–40% of familial ALS and 7% of sporadic form of ALS (Al-Chalabi et al., 2012). In FTD, pathogenic repeat expansion in *C9ORF72* accounts for 20–30% of familial form and about 6% of sporadic form of FTD (DeJesus-Hernandez et al., 2011), and is the most common cause of disease besides pathogenic mutations in *GRN* and *MAPT* (Sirkis et al., 2019). The *C9ORF72* repeat also showed segregation with AD status in three families (Harms et al., 2013). In addition to the *C9ORF72* repeat, the number of CA repeat in *STMN2* (Theunissen et al., 2021) and CAG repeat in *ATXN2* (Elden et al., 2010) were associated with risk and age-of-onset of ALS.

In PD, mutations in 18 *PARK* genes have been found as the main cause of familial PD (Klein and Westenberger, 2012). For example, exonic deletions in *Parkin* gene were the most common mutations in families with autosomal recessive parkinsonism (Taghavi et al., 2018), and SCNA duplication can lead to autosomal dominant PD (Singleton et al., 2003; Konno et al., 2016). Besides, STRs in a few ataxia genes can be a rare cause of PD. CAG repeat expansion in SCA2 gene *ATXN2* is also the cause of autosomal dominant PD (Table 2; Charles et al., 2007). The biallelic AAGGG repeat expansion in *RFC1* is a common cause of late-onset ataxia (Cortese et al., 2019) and is likely to be a rare cause of PD (Kytövuori et al., 2022). Repeat expansions in SCA3, SCA10, and SCA17 have been described in patients with PD as well (Park et al., 2015;

Schüle et al., 2017). In addition, the GGGGCC repeat expansion in *C9ORF72* is also a rare cause of PD (Table 2; Lesage et al., 2013).

As a type of minisatellite with sequence repeats vary between individuals, variable number of tandem repeats (VNTRs) are usually considered as longer cousins of STRs. A 69 bp VNTR in *WDR7* was found to be enriched in sporadic ALS patients (Table 2; Course et al., 2020). VNTR in *IL1RN* intron 2 was related to earlier Wilson's disease onset, particularly among patients with neuropsychiatric form of the disease (Table 2; Gromadzka and Członkowska, 2011). Frequency of 2R allele of *XRCC5* gene was significant different between multiple sclerosis (MS) patients and controls (Table 2; Jahantigh et al., 2017). Individuals with homozygous *INS* class III alleles (characterized by 141–209 repeats of 14 bp motif) showed earlier onset of AD (Table 2; Majores et al., 2002). An 25 bp intronic VNTR expansion in *ABCA7* was associated with GWAS SNP (rs3764650) and enriched in AD patients (Table 2, odds ratio = 4.5 (1.3–24.2); De Roeck et al., 2018).

The traditional method to detect STRs/VNTRs is low throughput as it involves performing polymerase chain reaction on target regions and gel electrophoresis (DeJesus-Hernandez et al., 2011). With the availability of sequencing data, detecting STRs/VNTRs in a large scale is practical. Currently, various algorithms have been developed for STR/VNTR detection from sequencing data, e.g., popSTR (Kristmundsdóttir et al., 2017), GangSTR (Mousavi et al., 2019), STRetch (Dashnow et al., 2018), and ExpansionHunter (Dolzhenko et al., 2019) for STR detection, and VNTRseek (Gelfand et al., 2014), adVNTR (Bakhtiari et al., 2018) and code-adVNTR (Park et al., 2022) for VNTR detection. A few reviews (Gymrek, 2017; Halman and Oshlack, 2020) have provided a comprehensive list of tools available for STR detection from sequencing data. Considering the successful application of profiling STRs/VNTRs associated with disease status and gene expression (Bakhtiari et al., 2018; Fotsing et al., 2019; Mousavi et al., 2019; van der Sanden et al., 2021), studying STRs/VNTRs in a large cohort of AD patients may yield new discovery to AD genetics. A list of neurodegenerative diseases caused by STRs/VNTRs are displayed in Table 2.

Transposable elements related SVs

Transposable elements (TEs) are DNA sequences that can move between different genomic positions. About 44% of the human genome is consisted of transposons: 20% long interspersed element (LINEs), 13% short interspersed elements (SINEs), 8% long terminal repeats (LTR) retro-transposons, and 3% DNA transposons (Mills et al., 2007). More importantly, TEs can serve as an active mutagen in the human genome, causing insertions, deletions, duplications, inversions, and translocations (Beck et al., 2011). It has been reported that TEs are activated in AD, causing genomic instability (Guo et al., 2018). TEs are also widely involved in a range of other neurodegenerative diseases (Tam et al., 2019), implicating potential pathogenic transposon-induced SVs underlying the genetic basis of neurodegenerative disorders.

One class of TEs, called SINE-VNTR-Alu (SVA), is of particular interest since it naturally harbors VNTRs with gene-regulatory functions inside its body. It was reported that 82% of structurally variable SVAs were around 50 Kb of a transcription start site, indicating their role as a transcriptional regulatory element in

disease pathogenesis (van Bree et al., 2022). In the same study, a few SVAs in the LD blocks of GWAS SNP signals for AD, PD, MS, and ALS were identified; the deletion of SVAs around *BINI* (AD gene) and *BCKDK* (PD gene) alters epigenome and nearby gene expression (Table 2; van Bree et al., 2022). X-linked dystonia-parkinsonism (XDP) patients carry a 294 Kb identical founder haplotype harboring a few sequence variants. Particularly, the number of CCCTCT repeat inside the SVA element in this region displayed highly significant inverse correlation with age of XDP onset, suggesting this repeat could be the causative variant (Table 2; Bragg et al., 2017). In a study of SVAs in PD, no significant association with disease risk was reported; however, one SVA inside the H1 haplotype on chromosome 17 was reported and found to be associated with the expression of multiple genes at this locus (Pfaff et al., 2021).

In AD, there is evidence that primate-specific Alu retrotransposons repeatedly inserted into the intron of *TOMM40* (Table 2; Larsen et al., 2017), which has been implicated in the pathogenesis of AD by numerous GWAS studies. The deoxythymidine homopolymer repeat (rs10524523), a part of Alu mobile element monomer, is inserted into the intron 6 of *TOMM40*, and is usually in the same haplotypes as *APOE* alleles with long repeats (18 bp <and <30 bp) corresponding to *APOE* E4 and short repeats (≤ 18 bp) or very long repeats (≥ 30 bp) corresponding to *APOE* E2/E3 in Caucasians (Crenshaw et al., 2013). A list of neurodegenerative diseases caused by TE related SVs were displayed in Table 2.

Discussion

From GWASs in the past decade, dozens of SNPs with small effects were associated with AD. Among of those identified SNPs, *APOE4/2* remained as the most powerful alleles in predicting the risk of LOAD. *APOE* alleles alone can reach an AUC (Area Under the Curve) of 0.70 in predicting AD, while the best AUC is 0.61 for all other SNPs combined (Leonenko et al., 2021). SVs are major genetic mutations residing in the genome besides SNPs. Extending current genetical analysis further into the field of SVs would definitely help improve the polygenic prediction of AD risk, therefore, facilitating earlier diagnosis and treatment before irreversible pathological damages were done. Moreover, it was pointed out that the most important limiting factor in the translation of knowledge from genetics to drugs is the lack of good models for AD (Sierksma et al., 2020). Potential mechanism of actions in disease etiology identified by SVs might lead to new animal models and drug targets for the treatment of AD.

So far, the *APP* duplication is the only structural variation with sufficient supports for directly causing AD. Other than that, a few candidate SVs (Table 1) associated with increased risk of AD are identified. The reasons for the lack of new discoveries are multifold. First, only large CNVs can be detected on SNP array. Those large CNVs tend to be rare and are not likely to reach statistical significance without sufficient sample size. Second, even though smaller SVs can be obtained from short-read sequencing data, reliable read mapping for SV calling from short reads still remains challenging since half of the human genome is made of repetitive sequences (Lander et al., 2001). Third, unlike SNPs which affect single locus, SVs are usually represented by a range that may have different breakpoints in different individuals, therefore, increasing the difficulties in analyzing SVs.

In the future, with the revolution and refinement in sequencing techniques, particularly, long sequencing techniques, we believe that we would be able to characterize SVs in a large sample size accurately, exhaustively, and cost-efficiently. The issue of low mappability across a range of “problematic” genomic locations can be overcome. In this way, the foundation for discovering disease related SVs would be laid. At the same time, short-read sequencing can help correct the higher error rate of long read sequencing on each individual nucleotide to facilitate better SNP and small indel calling. Finally, with the incorporation of multi-omics data/single-cell data, researchers are able to fill the gap between genotype and phenotype, therefore, uncovering the full landscape of disease etiology.

Author contributions

W-PL and L-SW conceived the idea. HW and W-PL carried out the structure of the writing and contributed to manuscript writing. GS supervised the project. All authors contributed to the article and approved the submitted version.

References

- Abel, H. J., Larson, D. E., Regier, A. A., Chiang, C., Das, I., Kanchi, K. L., et al. (2020). Mapping and characterization of structural variation in 17,795 human genomes. *Nature* 583, 83–89. doi: 10.1038/s41586-020-2371-0
- Al Khleifat, A., Iacoangeli, A., Van Vugt, J. J., Bowles, H., Moisse, M., Zwamborn, R. A., et al. (2022). Structural variation analysis of 6,500 whole genome sequences in amyotrophic lateral sclerosis. *NPJ Genom. Med.* 7:8. doi: 10.1038/s41525-021-00267-9
- Al-Chalabi, A., Jones, A., Troakes, C., King, A., Al-Sarraj, S., and Van Den Berg, L. H. (2012). The genetics and neuropathology of amyotrophic lateral sclerosis. *Acta Neuropathol.* 124, 339–352. doi: 10.1007/s00401-012-1022-4
- Allen, M., Kachadoorian, M., Quicksall, Z., Zou, F., Chai, H. S., Younkin, C., et al. (2014). Association of MAPT haplotypes with Alzheimer’s disease risk and MAPT brain gene expression levels. *Alzheimers Res. Ther.* 6:39. doi: 10.1186/alzrt268
- Anheim, M., Mariani, L.-L., Calvas, P., Cheuret, E., Zagnoli, F., Odent, S., et al. (2012). Exonic deletions of FXN and early-onset Friedreich ataxia. *Arch. Neurol.* 69, 912–916. doi: 10.1001/archneur.2011.834
- Arai, T., Ikeda, K., Akiyama, H., Shikamoto, Y., Tsuchiya, K., Yagishita, S., et al. (2001). Distinct isoforms of tau aggregated in neurons and glial cells in brains of patients with Pick’s disease, corticobasal degeneration and progressive supranuclear palsy. *Acta Neuropathol.* 101, 167–173. doi: 10.1007/s004010000283
- Baker, M., Litvan, I., Houlden, H., Adamson, J., Dickson, D., Perez-Tur, J., et al. (1999). Association of an extended haplotype in the tau gene with progressive supranuclear palsy. *Hum. Mol. Genet.* 8, 711–715. doi: 10.1093/hmg/8.4.711
- Bakhtiar, M., Shleizer-Burko, S., Gymrek, M., Bansal, V., and Bafna, V. (2018). Targeted genotyping of variable number tandem repeats with aVNTR. *Genome Res.* 28, 1709–1719. doi: 10.1101/gr.235119.118
- Banfi, S., Servadio, A., Chung, M., Kwiatkowski, T. J., McCall, A. E., Duvick, L. A., et al. (1994). Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nat. Genet.* 7, 513–520. doi: 10.1038/ng0894-513
- Beck, C. R., Garcia-Perez, J. L., Badge, R. M., and Moran, J. V. (2011). LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.* 12, 187–215. doi: 10.1146/annurev-genom-082509-141802
- Bejjani, B. A., and Shaffer, L. G. (2006). Application of array-based comparative genomic hybridization to clinical diagnostics. *J. Mol. Diagn.* 8, 528–533. doi: 10.2353/jmoldx.2006.060029
- Bellenguez, C., Küçükali, F., Jansen, I. E., Kleindam, L., Moreno-Grau, S., Amin, N., et al. (2022). New insights into the genetic etiology of Alzheimer’s disease and related dementias. *Nat. Genet.* 54, 412–436. doi: 10.1038/s41588-022-01024-z
- Bertram, L., Lill, C. M., and Tanzi, R. E. (2010). The genetics of Alzheimer disease: back to the future. *Neuron* 68, 270–281. doi: 10.1016/j.neuron.2010.10.013
- Bhidayasiri, R., Perlman, S. L., Pulst, S.-M., and Geschwind, D. H. (2005). Late-onset Friedreich ataxia: phenotypic analysis, magnetic resonance imaging findings, and review of the literature. *Arch. Neurol.* 62, 1865–1869. doi: 10.1001/archneur.62.12.1865
- Blom, E. S., Viswanathan, J., Kilander, L., Helisalmi, S., Soyninen, H., Lannfelt, L., et al. (2008). Low prevalence of APP duplications in Swedish and Finnish patients with early-onset Alzheimer’s disease. *Eur. J. Hum. Genet.* 16, 171–175. doi: 10.1038/sj.ejhg.5201966
- Boettger, L. M., Handsaker, R. E., Zody, M. C., and McCarroll, S. A. (2012). Structural haplotypes and recent evolution of the human 17q21.31 region. *Nat. Genet.* 44, 881–885. doi: 10.1038/ng.2334
- Bragg, D. C., Mangkalaphiban, K., Vaine, C. A., Kulkarni, N. J., Shin, D., Yadav, R., et al. (2017). Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in TAF1. *Proc. Natl. Acad. Sci. U. S. A.* 114, E11020–E11028. doi: 10.1073/pnas.1712526114
- Brouwers, N., Van Cauwenbergh, C., Engelborghs, S., Lambert, J. C., Bettens, K., Le Bastard, N., et al. (2012). Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites. *Mol. Psychiatry* 17, 223–233. doi: 10.1038/mp.2011.24
- Buizer-Voskamp, J. E., Muntjewerff, J.-W., Risk, G., Strengman, E., Sabatti, C., Stefansson, H., et al. (2011). Genome-wide analysis shows increased frequency of copy number variation deletions in Dutch schizophrenia patients. *Biol. Psychiatry* 70, 655–662. doi: 10.1016/j.biopsych.2011.02.015
- Bürk, K., Abele, M., Fetter, M., Dichgans, J., Skalej, M., Laccione, F., et al. (1996). Autosomal dominant cerebellar ataxia type I clinical features and MRI in families with SCA1, SCA2 and SCA3. *Brain* 119, 1497–1505. doi: 10.1093/brain/119.5.1497
- Butcher, N. J., Horne, M. K., Mellick, G. D., Fowler, C. J., Masters, C. L., and Minchin, R. F. (2018). Sulfotransferase 1A3/4 copy number variation is associated with neurodegenerative disease. *Pharm. J.* 18, 209–214. doi: 10.1038/tpj.2017.4
- Byman, E., Nägga, K., Gustavsson, A.-M., Andersson-Assarsson, J., Hansson, O., Sonestedt, E., et al. (2020). Alpha-amylase 1A copy number variants and the association with memory performance and Alzheimer’s dementia. *Alzheimers Res. Ther.* 12, 1–10. doi: 10.1186/s13195-020-00726-y
- Byman, E., Schultz, N., Bank, N., Bfex, M., and Wennström, M. (2018). Brain alpha-amylase: a novel energy regulator important in Alzheimer disease? *Brain Pathol.* 28, 920–932. doi: 10.1111/bpa.12597
- Cáceres, A., Sindi, S. S., Raphael, B. J., Cáceres, M., and González, J. R. (2012). Identification of polymorphic inversions from genotypes. *BMC Bioinformatics* 13:28. doi: 10.1186/1471-2105-13-28
- Caffrey, T. M., Joachim, C., Paracchini, S., Esiri, M. M., and Wade-Martins, R. (2006). Haplotype-specific expression of exon 10 at the human MAPT locus. *Hum. Mol. Genet.* 15, 3529–3537. doi: 10.1093/hmg/ddl429
- Campuzano, V., Montermini, L., Molto, M. D., Pianese, L., Cossée, M., Cavalcanti, F., et al. (1996). Friedreich’s ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271, 1423–1427. doi: 10.1126/science.271.5254.1423
- Chan, S., Lam, E., Saghbini, M., Bocklandt, S., Hastie, A., Cao, H., et al. (2018). “Structural variation detection and analysis using bionano optical mapping” in *Copy Number Variants: Methods in Molecular Biology*. Vol. 1833. ed. D. Bickhart (New York, NY: Humana Press). 193–203.
- Charles, P., Camuzat, A., Benammar, N., Sellal, F., Destee, A., Bonnet, A. M., et al. (2007). Are interrupted SCA2 CAG repeat expansions responsible for parkinsonism? *Neurology* 69, 1970–1975. doi: 10.1212/01.wnl.0000269323.21969.db
- Chen, Z., Chen, J. A., Shatunov, A., Jones, A. R., Kravitz, S. N., Huang, A. Y., et al. (2019). Genome-wide survey of copy number variants finds MAPT duplications in progressive supranuclear palsy. *Mov. Disord.* 34, 1049–1059. doi: 10.1002/mds.27702

Funding

This work was supported by funding from the National Institute on Aging (RF1AG074328 and U24-AG041689) and University of Pennsylvania ADRC pilot project (R01-AG064877-02S1).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., et al. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32, 1220–1222. doi: 10.1093/bioinformatics/btv710
- Cheng, H.-L., Shao, Y.-R., Dong, Y., Dong, H.-L., Yang, L., Ma, Y., et al. (2021). Genetic spectrum and clinical features in a cohort of Chinese patients with autosomal recessive cerebellar ataxias. *Transl. Neurodegener.* 10, 1–15. doi: 10.1186/s40035-021-00264-z
- Colella, S., Yau, C., Taylor, J. M., Mirza, G., Butler, H., Clouston, P., et al. (2007). QuantiSNP: an objective Bayes hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* 35, 2013–2025. doi: 10.1093/nar/gkm076
- Cortese, A., Simone, R., Sullivan, R., Vandrovicova, J., Tariq, H., Yau, W. Y., et al. (2019). Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat. Genet.* 51, 649–658. doi: 10.1038/s41588-019-0372-4
- Course, M. M., Gudsnuk, K., Smukowski, S. N., Winston, K., Desai, N., Ross, J. P., et al. (2020). Evolution of a human-specific tandem repeat associated with ALS. *Am. J. Hum. Genet.* 107, 445–460. doi: 10.1016/j.ajhg.2020.07.004
- Crenshaw, D. G., Gottschalk, W. K., Lutz, M. W., Grossman, I., Saunders, A. M., Burke, J. R., et al. (2013). Using genetics to enable studies on the prevention of Alzheimer's disease. *Clin. Pharmacol. Ther.* 93, 177–185. doi: 10.1038/clpt.2012.222
- Crook, R., Verkoniemi, A., Perez-Tur, J., Mehta, N., Baker, M., Houlden, H., et al. (1998). A variant of Alzheimer's disease with spastic paraparesis and unusual plaques due to deletion of exon 9 of presenilin 1. *Nat. Med.* 4, 452–455. doi: 10.1038/nm0498-452
- Damji, K. F., Allingham, R. R., Pollock, S. C., Small, K., Lewis, K. E., Stajich, J. M., et al. (1996). Periodic vestibulocerebellar ataxia, an autosomal dominant ataxia with defective smooth pursuit, is genetically distinct from other autosomal dominant ataxias. *Arch. Neurol.* 53, 338–344. doi: 10.1001/archneur.1996.00550040074016
- Dashnow, H., Lek, M., Phipson, B., Halman, A., Sadedin, S., Lonsdale, A., et al. (2018). STRetch: detecting and discovering pathogenic short tandem repeat expansions. *Genome Biology* 9, 1–13. doi: 10.1186/s13059-018-1505-2
- De Coster, W., and Van Broeckhoven, C. (2019). Newest methods for detecting structural variations. *Trends Biotechnol.* 37, 973–982. doi: 10.1016/j.tibtech.2019.02.003
- De Roeck, A., Duchateau, L., Van Dongen, J., Cacace, R., Bjerke, M., Van den Bossche, T., et al. (2018). An intronic VNTR affects splicing of ABCA7 and increases risk of Alzheimer's disease. *Acta Neuropathol.* 135, 827–837. doi: 10.1007/s00401-018-1841-z
- DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245–256. doi: 10.1016/j.neuron.2011.09.011
- Dolzhenko, E., Deshpande, V., Schlesinger, F., Krusche, P., Petrovski, R., Chen, S., et al. (2019). ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* 35, 4754–4756. doi: 10.1093/bioinformatics/btz431
- Ebbert, M. T. W., Jensen, T. D., Jansen-West, K., Sens, J. P., Reddy, J. S., Ridge, P. G., et al. (2019). Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biol.* 20:97. doi: 10.1186/s13059-019-1707-2
- Elden, A. C., Kim, H.-J., Hart, M. P., Chen-Plotkin, A. S., Johnson, B. S., Fang, X., et al. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 466, 1069–1075. doi: 10.1038/nature09320
- Erdogan, F., Chen, W., Kirchhoff, M., Kalscheuer, V. M., Hultschig, C., Müller, I., et al. (2006). Impact of low copy repeats on the generation of balanced and unbalanced chromosomal aberrations in mental retardation. *Cytogenet. Genome Res.* 115, 247–253. doi: 10.1159/000095921
- Falconer, E., Hills, M., Naumann, U., Poon, S. S., Chavez, E. A., Sanders, A. D., et al. (2012). DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. *Nat. Methods* 9, 1107–1112. doi: 10.1038/nmeth.2206
- Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97. doi: 10.1038/nrg1767
- Fotsing, S. F., Margoliash, J., Wang, C., Saini, S., Yanicky, R., Shleizer-Burko, S., et al. (2019). The impact of short tandem repeat variation on gene expression. *Nat. Genet.* 51, 1652–1659. doi: 10.1038/s41588-019-0521-9
- Gatz, M., Reynolds, C. A., Frattiglioni, L., Johansson, B., Mortimer, J. A., Berg, S., et al. (2006). Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* 63, 168–174. doi: 10.1001/archpsyc.63.2.168
- Gelfand, Y., Hernandez, Y., Loving, J., and Benson, G. (2014). VNTRseek—a computational tool to detect tandem repeat variants in high-throughput sequencing data. *Nucleic Acids Res.* 42, 8884–8894. doi: 10.1093/nar/gku642
- Ghani, M., Pinto, D., Lee, J. H., Grinberg, Y., Sato, C., Moreno, D., et al. (2012). Genome-wide survey of large rare copy number variants in Alzheimer's disease among Caribbean hispanics. *G3* 2, 71–78. doi: 10.1534/g3.111.000869
- Gromadzka, G., and Czlonkowska, A. (2011). Influence of IL-1RN intron 2 variable number of tandem repeats (VNTR) polymorphism on the age at onset of neuropsychiatric symptoms in Wilson's disease. *Int. J. Neurosci.* 121, 8–15. doi: 10.3109/00207454.2010.523131
- Guan, P., and Sung, W.-K. (2016). Structural variation detection using next-generation sequencing data: a comparative technical review. *Methods* 102, 36–49. doi: 10.1016/j.ymeth.2016.01.020
- Guffanti, G., Torri, F., Rasmussen, J., Clark, A. P., Lakatos, A., Turner, J. A., et al. (2013). Increased CNV-region deletions in mild cognitive impairment (MCI) and Alzheimer's disease (AD) subjects in the ADNI sample. *Genomics* 102, 112–122. doi: 10.1016/j.ygeno.2013.04.004
- Guo, C., Jeong, H.-H., Hsieh, Y.-C., Klein, H.-U., Bennett, D. A., De Jager, P. L., et al. (2018). Tau activates transposable elements in Alzheimer's disease. *Cell Rep.* 23, 2874–2880. doi: 10.1016/j.celrep.2018.05.004
- Gymrek, M. (2017). A genomic view of short tandem repeats. *Curr. Opin. Gen. Dev.* 44, 9–16. doi: 10.1016/j.celrep.2017.01.012
- Halman, A., and Oshlack, A. (2020). Accuracy of short tandem repeats genotyping tools in whole exome sequencing data. *F1000Research* 9. doi: 10.12688/f1000research.22639.1
- Hardy, J., Pittman, A., Myers, A., Gwinn-Hardy, K., Fung, H. C., De Silva, R., et al. (2005). Evidence suggesting that homo neanderthalensis contributed the H2 MAPT haplotype to *Homo sapiens*. *Biochem. Soc. Trans.* 33, 582–585. doi: 10.1042/BST0330582
- Harms, M., Benitez, B. A., Cairns, N., Cooper, B., Cooper, P., Mayo, K., et al. (2013). C9orf72 hexanucleotide repeat expansions in clinical Alzheimer disease. *JAMA Neurol.* 70, 736–741. doi: 10.1001/2013.jamaneurol.537
- Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M. L., et al. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* 41, 1088–1093. doi: 10.1038/ng.440
- Heinzen, E. L., Need, A. C., Hayden, K. M., Chiba-Falek, O., Roses, A. D., Strittmatter, W. J., et al. (2010). Genome-wide scan of copy number variation in late-onset Alzheimer's disease. *J. Alzheimers Dis.* 19, 69–77. doi: 10.3233/JAD-2010-1212
- Ho, S. S., Urban, A. E., and Mills, R. E. (2020). Structural variation in the sequencing era. *Nat. Rev. Genet.* 21, 171–189. doi: 10.1038/s41576-019-0180-9
- Hooli, B. V., Kovacs-Vajna, Z. M., Mullin, K., Blumenthal, M. A., Mattheisen, M., Zhang, C., et al. (2014). Rare autosomal copy number variations in early-onset familial Alzheimer's disease. *Mol. Psychiatry* 19, 676–681. doi: 10.1038/mp.2013.77
- Hooli, B. V., Mohapatra, G., Mattheisen, M., Parrado, A. R., Roeher, J. T., Shen, Y., et al. (2012). Role of common and rare APP DNA sequence variants in Alzheimer disease. *Neurology* 78, 1250–1257. doi: 10.1212/WNL.0b013e3182515972
- Iourov, I. Y., Vorsanova, S. G., Liehr, T., and Yurov, Y. B. (2009). Aneuploidy in the normal, Alzheimer's disease and ataxia-telangiectasia brain: differential expression and pathological meaning. *Neurobiol. Dis.* 34, 212–220. doi: 10.1016/j.nbd.2009.01.003
- Jacobs, P., Baikie, A. G., and Strong, J. A. (1959). The somatic chromosomes in mongolism. *Lancet* 273:710. doi: 10.1016/S0140-6736(59)91892-6
- Jahantigh, D., Moghtaderi, A., Naroie-Nejad, M., Mousavi, M., Moosavi, M., Salimi, S., et al. (2017). Carriage of 2R allele at VNTR polymorphous site of XRCC5 gene increases risk of multiple sclerosis in an Iranian population. *Russ. J. Genet.* 53, 147–152. doi: 10.1134/S102279541612005X
- Jain, M., Olsen, H. E., Turner, D. J., Stoddart, D., Bulazel, K. V., Paten, B., et al. (2018). Linear assembly of a human centromere on the Y chromosome. *Nat. Biotechnol.* 36, 321–323. doi: 10.1038/nbt.4109
- Kallioniemi, A., Kallioniemi, O.-P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., et al. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258, 818–821. doi: 10.1126/science.1359641
- Kasuga, K., Shimohata, T., Nishimura, A., Shiga, A., Mizuguchi, T., Tokunaga, J., et al. (2009). Identification of independent APP locus duplication in Japanese patients with early-onset Alzheimer disease. *J. Neurol. Neurosurg. Psychiatry* 80, 1050–1052. doi: 10.1136/jnnp.2008.161703
- Kim, M.-O., Takada, L. T., Wong, K., Forner, S. A., and Geschwind, M. D. (2018). Genetic PRP prion diseases. *Cold Spring Harb. Perspect. Biol.* 10:a033134. doi: 10.1101/cshperspect.a033134
- Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K. K., Holmans, P., et al. (2009a). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum. Mol. Genet.* 18, 1497–1503. doi: 10.1093/hmg/ddp043
- Kirov, G., Rujescu, D., Ingason, A., Collier, D. A., O'Donovan, M. C., and Owen, M. J. (2009b). Neurexin 1 (NRXN1) deletions in schizophrenia. *Schizophr. Bull.* 35, 851–854. doi: 10.1093/schbul/sbp079
- Klein, C., and Westenberger, A. (2012). Genetics of Parkinson's disease. *Cold Spring Harb. Perspect. Med.* 2:a008888. doi: 10.1101/cshperspect.a008888
- Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., et al. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidolusian atrophy (DRPLA). *Nat. Genet.* 6, 9–13. doi: 10.1038/ng0194-9
- Konno, T., Ross, O. A., Puschmann, A., Dickson, D. W., and Wszolek, Z. K. (2016). Autosomal dominant Parkinson's disease caused by SNCA duplications. *Parkinsonism Relat. Disord.* 22, S1–S6. doi: 10.1016/j.parkreidis.2015.09.007
- Koolen, D. A., Vissers, L. E., Pfundt, R., De Leeuw, N., Knight, S. J., Regan, R., et al. (2006). A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat. Genet.* 38, 999–1001. doi: 10.1038/ng1853
- Kremer, E. J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., et al. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p (CCG)n. *Science* 252, 1711–1714. doi: 10.1126/science.1675488

- Kristmundsdóttir, S., Sigurpáldóttir, B. D., Kehr, B., and Halldórsson, B. V. (2017). popSTR: population-scale detection of STR variants. *Bioinformatics* 33, 4041–4048. doi: 10.1093/bioinformatics/btw568
- Kushima, I., Aleksic, B., Nakatochi, M., Shimamura, T., Okada, T., Uno, Y., et al. (2018). Comparative analyses of copy-number variation in autism spectrum disorder and schizophrenia reveal etiological overlap and biological insights. *Cell Rep.* 24, 2838–2856. doi: 10.1016/j.celrep.2018.08.022
- Kytövuori, L., Sipilä, J., Hurme-Niiranen, A., Siitonen, A., Koshimizu, E., Miyatake, S., et al. (2022). Biallelic expansion in RFC1 as a rare cause of Parkinson's disease. *NPJ Parkinsons Dis.* 8, 1–4. doi: 10.1038/s41531-021-00275-7
- Lambert, J.-C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., et al. (2009). Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet.* 41, 1094–1099. doi: 10.1038/ng.439
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Larsen, P. A., Lutz, M. W., Hunnicutt, K. E., Mihovilovic, M., Saunders, A. M., Yoder, A. D., et al. (2017). The Alu neurodegeneration hypothesis: a primate-specific mechanism for neuronal transcription noise, mitochondrial dysfunction, and manifestation of neurodegenerative disease. *Alzheimers Dement.* 13, 828–838. doi: 10.1016/j.jalz.2017.01.017
- Layer, R. M., Chiang, C., Quinlan, A. R., and Hall, I. M. (2014). LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* 15, 1–19. doi: 10.1186/gb-2014-15-6-r84
- Lee, S. H., Harold, D., Nyholt, D. R., ANZGene ConsortiumInternational Endogene ConsortiumGenetic and Environmental Risk for Alzheimer's disease Consortium et al. (2013). Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum. Mol. Genet.* 22, 832–841. doi: 10.1093/hmg/ddt491
- Lee, W.-P., Tucci, A. A., Conery, M., Leung, Y. Y., Kuzma, A. B., Valladares, O., et al. (2021). Copy number variation identification on 3,800 Alzheimer's disease whole genome sequencing data from the Alzheimer's disease sequencing project. *Front. Genet.* 12:752390. doi: 10.3389/fgene.2021.752390
- Leonenko, G., Baker, E., Stevenson-Hoare, J., Sierksma, A., Fiers, M., Williams, J., et al. (2021). Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. *Nat. Commun.* 12, 1–10. doi: 10.1038/s41467-021-24082-z
- Lesage, S., Le Ber, I., Condroyer, C., Broussolle, E., Gabelle, A., Thobois, S., et al. (2013). C9orf72 repeat expansions are a rare genetic cause of parkinsonism. *Brain* 136, 385–391. doi: 10.1093/brain/awt357
- Levy-Lahad, E., Wasco, W., Poorkaj, P., Romano, D. M., Oshima, J., Pettingell, W. H., et al. (1995). Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* 269, 973–977. doi: 10.1126/science.7638622
- Li, Y., Shaw, C. A., Sheffer, I., Sule, N., Powell, S. Z., Dawson, B., et al. (2012). Integrated copy number and gene expression analysis detects a CREB1 association with Alzheimer's disease. *Transl. Psychiatry* 2:e192. doi: 10.1038/tp.2012.119
- Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369
- Lin, K., Smit, S., Bonnema, G., Sanchez-Perez, G., and de Ridder, D. (2015). Making the difference: integrating structural variation detection tools. *Brief. Bioinform.* 16, 852–864. doi: 10.1093/bib/bbu047
- Majores, M., Kölsch, H., Bagli, M., Ptok, U., Kockler, M., Becker, K., et al. (2002). The insulin gene VNTR polymorphism in Alzheimer's disease: results of a pilot study. *J. Neural Transm.* 109, 1029–1034. doi: 10.1007/s007020200086
- Manning, M., Hudgins, L., Practice, P., and Committee, G. (2010). Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet. Med.* 12, 742–745. doi: 10.1097/GIM.0b013e3181f8baad
- Maury, E. A., Sherman, M. A., Genovese, G., Gilgenast, T. G., Rajarajan, P., Flaherty, E., et al. (2022). Schizophrenia-associated somatic copy number variants from 12,834 cases reveal contribution to risk and recurrent, isoform-specific NRXN1 disruptions. *medRxiv* 2021–12.
- McCoy, R. C., Taylor, R. W., Blauwkamp, T. A., Kelley, J. L., Kertesz, M., Pushkarev, D., et al. (2014). Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. *PLoS One* 9:e106689. doi: 10.1371/journal.pone.0106689
- Miller, D. T., Adam, M. P., Aradhya, S., Biasecker, L. G., Brothman, A. R., Carter, N. P., et al. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* 86, 749–764. doi: 10.1016/j.ajhg.2010.04.006
- Mills, R. E., Bennett, E. A., Iskow, R. C., and Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends Genet.* 23, 183–191. doi: 10.1016/j.tig.2007.02.006
- Ming, C., Wang, M., Wang, Q., Neff, R., Wang, E., Shen, Q., et al. (2021). Whole genome sequencing-based copy number variations reveal novel pathways and targets in Alzheimer's disease. *Alzheimers Dement.* 18, 1846–1867. doi: 10.1002/alz.12507
- Moore, R. A., Herzog, C., Errett, J., Kocisko, D. A., Arnold, K. M., Hayes, S. F., et al. (2006). Octapeptide repeat insertions increase the rate of protease-resistant prion protein formation. *Protein Sci.* 15, 609–619. doi: 10.1110/ps.051822606
- Moore, R. C., Xiang, F., Monaghan, J., Han, D., Zhang, Z., Edström, L., et al. (2001). Huntington disease phenocopy is a familial prion disease. *Am. J. Hum. Genet.* 69, 1385–1388. doi: 10.1086/324414
- Mosch, B., Morawski, M., Mittag, A., Lenz, D., Tarnok, A., and Arendt, T. (2007). Aneuploidy and DNA replication in the normal human brain and Alzheimer's disease. *J. Neurosci.* 27, 6859–6867. doi: 10.1523/JNEUROSCI.0379-07.2007
- Mousavi, N., Shleizer-Burko, S., Yanicky, R., and Gymrek, M. (2019). Profiling the genome-wide landscape of tandem repeat expansions. *Nucleic Acids Res.* 47:e90. doi: 10.1093/nar/gkz501
- Nakamura, K., Jeong, S.-Y., Uchihara, T., Anno, M., Nagashima, K., Nagashima, T., et al. (2001). SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum. Mol. Genet.* 10, 1441–1448. doi: 10.1093/hmg/10.14.1441
- Park, J., Bakhtiari, M., Popp, B., Wiesener, M., and Bafna, V. (2022). Detecting tandem repeat variants in coding regions using code-adVNTR. *IScience* 25:104785. doi: 10.1016/j.isci.2022.104785
- Park, H., Kim, H.-J., and Jeon, B. S. (2015). Parkinsonism in spinocerebellar ataxia. *Biomed. Res. Int.* 2015:125273. doi: 10.1155/2015/125273
- Penney, J. B. Jr., Vonsattel, J.-P., Macdonald, M. E., Gusella, J. F., and Myers, R. H. (1997). CAG repeat number governs the development rate of pathology in Huntington's disease. *Ann. Neurol.* 41, 689–692. doi: 10.1002/ana.410410521
- Pfaff, A. L., Bubb, V. J., Quinn, J. P., and Koks, S. (2021). Reference SVA insertion polymorphisms are associated with Parkinson's disease progression and differential gene expression. *NPJ Parkinsons Dis.* 7, 1–9. doi: 10.1038/s41531-021-00189-4
- Pinkel, D., Straume, T., and Gray, J. W. (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. U. S. A.* 83, 2934–2938. doi: 10.1073/pnas.83.9.2934
- Pittman, A. M., Fung, H.-C., and de Silva, R. (2006). Untangling the tau gene association with neurodegenerative disorders. *Hum. Mol. Genet.* 15, R188–R195. doi: 10.1093/hmg/ddl190
- Porubsky, D., Ebert, P., Audano, P. A., Vollger, M. R., Harvey, W. T., Marijon, P., et al. (2021). Fully phased human genome assembly without parental data using single-cell strand sequencing and long reads. *Nat. Biotechnol.* 39, 302–308. doi: 10.1038/s41587-020-0719-5
- Raghavan, N., and Tosto, G. (2017). Genetics of Alzheimer's disease: the importance of polygenic and epistatic components. *Curr. Neurol. Neurosci. Rep.* 17:78. doi: 10.1007/s11910-017-0787-1
- Ridge, P. G., Mukherjee, S., Crane, P. K., Kauwe, J. S. K., and Consortium, A. D. G. (2013). Alzheimer's disease: analyzing the missing heritability. *PLoS One* 8:e79771. doi: 10.1371/journal.pone.0079771
- Rossi, G., Giaccone, G., Giampaolo, L., Iussich, S., Puoti, G., Frigo, M., et al. (2000). Creutzfeldt-Jakob disease with a novel four extra-repeat insertional mutation in the PrP gene. *Neurology* 55, 405–410. doi: 10.1212/WNL.55.3.405
- Rovelet-Lecrux, A., Hannequin, D., Raux, G., Meur, N. L., Laquerrière, A., Vital, A., et al. (2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.* 38, 24–26. doi: 10.1038/ng1718
- Rovelet-Lecrux, A., Lecourtis, M., Thomas-Anterion, C., Le Ber, I., Brice, A., Frebourg, T., et al. (2009). Partial deletion of the MAPT gene: a novel mechanism of FTDP-17. *Hum. Mutat.* 30, E591–E602. doi: 10.1002/humu.20979
- Rovelet-Lecrux, A., Legallic, S., Wallon, D., Flaman, J.-M., Martinaud, O., Bombois, S., et al. (2012). A genome-wide study reveals rare CNVs exclusive to extreme phenotypes of Alzheimer disease. *Eur. J. Hum. Genet.* 20, 613–617. doi: 10.1038/ejhg.2011.225
- Ryan, C. P. (2019). Tandem repeat disorders. *Evol. Med. Public Health* 2019:17. doi: 10.1093/emph/eoz005
- Salehi, S., Kabeer, F., Ceglia, N., Andronescu, M., Williams, M. J., Campbell, K. R., et al. (2021). Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature* 595, 585–590. doi: 10.1038/s41586-021-03648-3
- Sato, N., Amino, T., Kobayashi, K., Asakawa, S., Ishiguro, T., Tsunemi, T., et al. (2009). Spinocerebellar ataxia type 31 is associated with “inserted” penta-nucleotide repeats containing (TGGAA)_n. *Am. J. Hum. Genet.* 85, 544–557. doi: 10.1016/j.ajhg.2009.09.019
- Schüle, B., McFarland, K. N., Lee, K., Tsai, Y.-C., Nguyen, K.-D., Sun, C., et al. (2017). Parkinson's disease associated with pure ATXN10 repeat expansion. *NPJ Parkinsons Dis.* 3, 1–7. doi: 10.1038/s41531-017-0029-x
- Shakkottai, V. G., and Fogel, B. L. (2013). Clinical neurogenetics: autosomal dominant spinocerebellar ataxia. *Neurol. Clin.* 31, 987–1007. doi: 10.1016/j.ncl.2013.04.006
- Sherrington, R., Rogaeva, E. I., Liang, Y. A., Rogaeva, E. A., Levesque, G., Ikeda, M., et al. (1995). Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* 375, 754–760. doi: 10.1038/375754a0
- Sierksma, A., Escott-Price, V., and De Strooper, B. (2020). Translating genetic risk of Alzheimer's disease into mechanistic insight and drug targets. *Science* 370, 61–66. doi: 10.1126/science.abb8575

- Singleton, A. B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., et al. (2003). α -Synuclein locus triplication causes Parkinson's disease. *Science* 302:841. doi: 10.1126/science.1090278
- Sirkis, D. W., Geier, E. G., Bonham, L. W., Karch, C. M., and Yokoyama, J. S. (2019). Recent advances in the genetics of frontotemporal dementia. *Curr. Genet. Med. Rep.* 7, 41–52. doi: 10.1007/s40142-019-0160-6
- Sleegers, K., Brouwers, N., Gijssels, I., Theuns, J., Goossens, D., Wauters, J., et al. (2006). APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy. *Brain* 129, 2977–2983. doi: 10.1093/brain/awl203
- Smith, M. J., Kwok, J. B., McLean, C. A., Kril, J. J., Broe, G. A., Nicholson, G. A., et al. (2001). Variable phenotype of Alzheimer's disease with spastic paraparesis. *Ann. Neurol.* 49, 125–129. doi: 10.1002/1531-8249(200101)49:1<125::AID-ANA21>3.0.CO;2-1
- Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Döhner, H., et al. (1997). Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosom. Cancer* 20, 399–407. doi: 10.1002/(SICI)1098-2264(199712)20:4<399::AID-GCC12>3.0.CO;2-1
- Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O. P., Ingason, A., Steinberg, S., et al. (2008). Large recurrent microdeletions associated with schizophrenia. *Nature* 455, 232–236. doi: 10.1038/nature07229
- Stevanin, G., Fujigasaki, H., Lebre, A.-S., Camuzat, A., Jeannequin, C., Dodé, C., et al. (2003). Huntington's disease-like phenotype due to trinucleotide repeat expansions in the TBP and JPH3 genes. *Brain* 126, 1599–1603. doi: 10.1093/brain/awg155
- Stevanin, G., Trotter, Y., Cancel, G., Dürr, A., David, G., Didierjean, O., et al. (1996). Screening for proteins with polyglutamine expansions in autosomal dominant cerebellar ataxias. *Hum. Mol. Genet.* 5, 1887–1892. doi: 10.1093/hmg/5.12.1887
- Swaminathan, S., Huentelman, M. J., Corneveaux, J. J., Myers, A. J., Faber, K. M., Foroud, T., et al. (2012a). Analysis of copy number variation in Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals. *PLoS One* 7:e50640. doi: 10.1371/journal.pone.0050640
- Swaminathan, S., Kim, S., Shen, L., Risacher, S. L., Foroud, T., Pankratz, N., et al. (2011). Genomic copy number analysis in Alzheimer's disease and mild cognitive impairment: an ADNI study. *Int. J. Alzheimers Dis.* 2011:729478. doi: 10.4061/2011/729478
- Swaminathan, S., Shen, L., Kim, S., Inlow, M., West, J. D., M Faber, K., et al. (2012b). Analysis of copy number variation in Alzheimer's disease: the NIALOAD/NCRAD family study. *Curr. Alzheimer Res.* 9, 801–814. doi: 10.2174/156720512802455331
- Szigeti, K., Lal, D., Li, Y., Doody, R. S., Wilhelmsen, K., Yan, L., et al. (2013). Genome-wide scan for copy number variation association with age at onset of Alzheimer's disease. *J. Alzheimers Dis.* 33, 517–523. doi: 10.3233/JAD-2012-121285
- Taghavi, S., Chaoui, R., Tafakhori, A., Azcona, L. J., Firouzabadi, S. G., Omrani, M. D., et al. (2018). A clinical and molecular genetic study of 50 families with autosomal recessive Parkinsonism revealed known and novel gene mutations. *Mol. Neurobiol.* 55, 3477–3489. doi: 10.1007/s12035-017-0535-1
- Tam, O. H., Ostrow, L. W., and Gale Hammell, M. (2019). Diseases of the nERVous system: retrotransposon activity in neurodegenerative disease. *Mob. DNA* 10, 1–14. doi: 10.1186/s13100-019-0176-1
- Tanzi, R. E., Gusella, J. F., Watkins, P. C., Bruns, G. A., St George-Hyslop, P., Van Keuren, M. L., et al. (1987). Amyloid β protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus. *Science* 235, 880–884. doi: 10.1126/science.2949367
- Theunissen, F., Anderton, R. S., Mastaglia, F. L., Flynn, L. L., Winter, S. J., James, I., et al. (2021). Novel STMN2 variant linked to amyotrophic lateral sclerosis risk and clinical phenotype. *Front. Aging Neurosci.* 13:658226. doi: 10.3389/fnagi.2021.658226
- van Bree, E. J., Guimarães, R. L., Lundberg, M., Blujdea, E. R., Rosenkrantz, J. L., White, F. T., et al. (2022). A hidden layer of structural variation in transposable elements reveals potential genetic modifiers in human disease-risk loci. *Genome Res.* 32, 656–670. doi: 10.1101/gr.275515.121
- van den Bos, H., Spierings, D. C., Taudt, A., Bakker, B., Porubský, D., Falconer, E., et al. (2016). Single-cell whole genome sequencing reveals no evidence for common aneuploidy in normal and Alzheimer's disease neurons. *Genome Biol.* 17, 1–9. doi: 10.1186/s13059-016-0976-2
- van der Sanden, B. P., Corominas, J., de Groot, M., Pennings, M., Meijer, R. P., Verbeek, N., et al. (2021). Systematic analysis of short tandem repeats in 38,095 exomes provides an additional diagnostic yield. *Genet. Med.* 23, 1569–1573. doi: 10.1038/s41436-021-01174-1
- Vialle, R. A., de Paiva Lopes, K., Bennett, D. A., Cray, J. F., and Raj, T. (2022). Integrating whole-genome sequencing with multi-omic data reveals the impact of structural variants on gene regulation in the human brain. *Nat. Neurosci.* 25, 504–514. doi: 10.1038/s41593-022-01031-7
- Vissers, L. E., de Vries, B. B., Osoegawa, K., Janssen, I. M., Feuth, T., Choy, C. O., et al. (2003). Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am. J. Hum. Genet.* 73, 1261–1270. doi: 10.1086/379977
- Vittori, A., Breda, C., Repici, M., Orth, M., Roos, R. A., Outeiro, T. F., et al. (2014). Copy-number variation of the neuronal glucose transporter gene SLC2A3 and age of onset in Huntington's disease. *Human Molecular Genetics* 23, 3129–3137. doi: 10.1093/hmg/ddu022
- Walsh, T., McClellan, J. M., McCarthy, S. E., Addington, A. M., Pierce, S. B., Cooper, G. M., et al. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320, 539–543. doi: 10.1126/science.1155174
- Wang, H., Bennett, D. A., De Jager, P. L., Zhang, Q.-Y., and Zhang, H.-Y. (2021). Genome-wide epistasis analysis for Alzheimer's disease and implications for genetic risk prediction. *Alzheimers Res. Ther.* 13, 1–13. doi: 10.1186/s13195-021-00794-8
- Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S. F., et al. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* 17, 1665–1674. doi: 10.1101/gr.6861907
- Wang, X., Luan, Y., and Yue, F. (2022). EagleC: a deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps. *Sci. Adv.* 8:eabn9215. doi: 10.1126/sciadv.abn9215
- Wang, H., Yang, J., Schneider, J. A., De Jager, P. L., Bennett, D. A., and Zhang, H.-Y. (2020). Genome-wide interaction analysis of pathological hallmarks in Alzheimer's disease. *Neurobiol. Aging* 93, 61–68. doi: 10.1016/j.neurobiolaging.2020.04.025
- Ward, B. E., Cook, R. H., Robinson, A., Austin, J. H., and Summitt, R. L. (1979). Increased aneuploidy in Alzheimer disease. *Am. J. Med. Genet.* 3, 137–144. doi: 10.1002/ajmg.1320030204
- Wernick, A. I., Walton, R. L., Soto-Beasley, A. I., Koga, S., Heckman, M. G., Valentino, R. R., et al. (2021). Frequency of spinocerebellar ataxia mutations in patients with multiple system atrophy. *Clin. Auton. Res.* 31, 117–125. doi: 10.1007/s10286-020-00759-1
- Winchester, L., Yau, C., and Ragoussis, J. (2009). Comparing CNV detection methods for SNP arrays. *Brief Funct. Genomic. Proteomic.* 8, 353–366. doi: 10.1093/bfpg/elp017
- Yang, Y., Geldmacher, D. S., and Herrup, K. (2001). DNA replication precedes neuronal cell death in Alzheimer's disease. *J. Neurosci.* 21, 2661–2668. doi: 10.1523/JNEUROSCI.21-08-02661.2001
- Zabetian, C. P., Hutter, C. M., Factor, S. A., Nutt, J. G., Higgins, D. S., Griffith, A., et al. (2007). Association analysis of MAPT H1 haplotype and subhaplotypes in Parkinson's disease. *Ann. Neurol.* 62, 137–144. doi: 10.1002/ana.21157
- Zhang, L., Dong, X., Lee, M., Maslov, A. Y., Wang, T., and Vijg, J. (2019). Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc. Natl. Acad. Sci. U. S. A.* 116, 9014–9019. doi: 10.1073/pnas.1902510116
- Zhang, X., Du, R., Li, S., Zhang, F., Jin, L., and Wang, H. (2014). Evaluation of copy number variation detection for a SNP array platform. *BMC Bioinformatics* 15:50. doi: 10.1186/1471-2105-15-50
- Zheng, X., Demirci, F. Y., Barmada, M. M., Richardson, G. A., Lopez, O. L., Sweet, R. A., et al. (2015). Genome-wide copy-number variation study of psychosis in Alzheimer's disease. *Transl. Psychiatry* 5:e574. doi: 10.1038/tp.2015.64
- Zheng, G. X., Lau, B. T., Schnell-Levin, M., Jarosz, M., Bell, J. M., Hindson, C. M., et al. (2016). Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* 34, 303–311. doi: 10.1038/nbt.3432
- Zody, M. C., Jiang, Z., Fung, H.-C., Antonacci, F., Hillier, L. W., Cardone, M. F., et al. (2008). Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nat. Genet.* 40, 1076–1083. doi: 10.1038/ng.193