J Neuropathol Exp Neurol
Vol. 79, No. 6, June 2020, pp. 568–584
doi: 10.1093/jnen/nlaa024

REVIEW ARTICLE

The *MUC6/AP2A2* Locus and Its Relevance to Alzheimer's Disease: A Review

Peter T. Nelson, MD, PhD, David W. Fardo, PhD, and Yuriko Katsumata, PhD

Abstract

We recently reported evidence of Alzheimer's disease (AD)-linked genetic variation within the mucin 6 (*MUC6*) gene on chromosome 11p, nearby the adaptor-related protein complex 2 subunit alpha 2 (*AP2A2*) gene. This locus has interesting features related to human genomics and clinical research. *MUC6* gene variants have been reported to potentially influence viral—including herpesvirus—immunity and the gut microbiome. Within the *MUC6* gene is a unique variable number of tandem repeat (VNTR) region. We discovered an association between *MUC6* VNTR repeat expansion and AD pathologic severity, particularly tau proteinopathy. Here, we review the relevant literature. The AD-linked VNTR polymorphism may also influence *AP2A2* gene expression. *AP2A2* encodes a polypeptide component of the adaptor protein complex, AP-2, which is involved in clathrin-coated vesicle function and was previously implicated in AD pathogenesis. To provide background information, we describe some key knowledge gaps in AD genetics research. The "missing/hidden heritability problem" of AD is highlighted. Extensive portions of the human genome, including the *MUC6* VNTR, have not been thoroughly evaluated due to limitations of existing high-throughput sequencing technology. We present and discuss additional data, along with cautionary considerations, relevant to the hypothesis that *MUC6* repeat expansion influences AD pathogenesis.

Key Words: Alzheimer's Disease Sequencing Project (ADSP), Amyloid, APOE, Copy number variation (CNV), Endocytosis, Genome-wide association study (GWAS), Whole-exome sequencing (WES).

From the Sanders-Brown Center on Aging (PTN, DWF, YK); Department of Pathology (PTN); and Department of Biostatistics (DWF, YK), University of Kentucky, Lexington, Kentucky

Send correspondence to: Peter T. Nelson, MD, PhD, Rm 311 Sanders-Brown Center on Aging, 800 S. Limestone Avenue, Lexington, KY 40536; E-mail: peter.nelson@uky.edu

Funding included grants P30 AG028383, R01 AG057187, R01 AG042475, R56 AG057191, and U01 AG016976 from the National Institute on Aging/National Institutes of Health (NIH). The Genotype-Tissue Expression (GTEx) project was supported by the Common Fund of the Office of the Director of the NIH, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. We are deeply grateful to all of the study participants, clinical workers, and researchers who helped make this article possible. See Supplementary Data for additional acknowledgments.

The authors have no duality or conflicts of interest to declare.

Supplementary Data can be found at academic.oup.com/jnen.

INTRODUCTION

This review article is about a putative Alzheimer's disease (AD)-associated genetic polymorphism within the mucin 6 (*MUC6*) gene, and <4000 base pairs (bp) away from the gene that encodes adaptor-related protein complex 2 subunit alpha 2 (*AP2A2*), on human chromosome 11p15.5 (1). The genomic locus of interest is a poorly annotated variable number of tandem repeat (VNTR) region, largely unreadable by conventional sequencing methods. Here, we review the topical literature on *MUC6*, the risk-associated polymorphic VNTR region, and *AP2A2*. We examine whether it is credible that a novel high-impact risk allele could be discovered, given that so many AD-linked genetic loci have already been identified. Evidence is provided in support of the hypothesis that *MUC6* VNTR repeat expansion is associated with AD risk. We also discuss reasons to be skeptical of the hypothesis. Relevant data are presented from the University of Kentucky AD Center (UK-ADC) autopsy cohort. Finally, critical knowledge gaps are highlighted.

Mucin 6 (*MUC6*)

The *MUC6* gene is located in a recombination prone region approximately 1 million bp from the telomere of human chromosome 11p (2, 3). *MUC6* is clustered along with 3 other mucin family genes (*MUC2*, *MUC5A*, and *MUC5B*) and *AP2A2*. *MUC6* and *AP2A2* are in close proximity, oriented in opposite directions—their 3′-untranslated regions nearly overlap (Fig. 1). The juxtaposition of *AP2A2* and *MUC6*, with possible sharing of 3′ regulatory sequence elements, is conserved on chromosomes of other vertebrates including frogs (*Xenopus tropicalis*).

Mucin genes encode proteins that become glycosylated and serve as gel-like substances to protect and lubricate epithelial surfaces (4). The MUC6 protein is expressed preferentially in the gastrointestinal (GI) tract, and also in genitourinary and pulmonary epithelia (5–9). Here, we will focus on the repetitive domains and gene expression regulation of *MUC6*, as may be relevant to the hypothesis that the *MUC6* VNTR is an AD risk allele.

A conspicuous feature of the mucin gene family is that long repetitive genetic sequences reside in exons, and are transcribed and translated into proteins (10). The repetitive DNA regions encode polypeptide motifs enriched with residues for O-linked glycosylation—proline, threonine, and serine amino acids, thus, they are termed PTS domains. These glycosylated

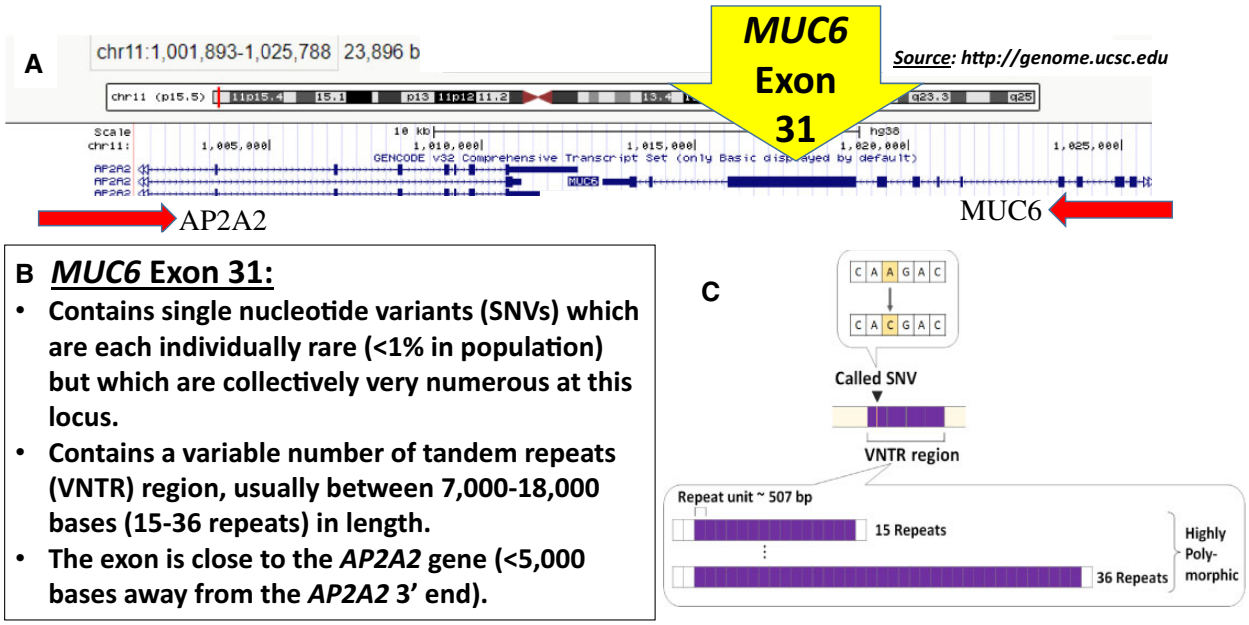

FIGURE 1. The region on human chromosome 11 which contains the *MUC6* variable number of tandem repeat (VNTR) region. This locus is ~1 million bp from the chromosome 11p telomere. The window of human chromosome 11p15.5 shown here **(A)** is relatively small (<24 kb) and contains only the 3′ portions of both *AP2A2* and *MUC6* genes. The yellow arrow indicates the *MUC6* exon 31 which contains the VNTR region. Note that, in this annotated assembly, the *MUC6* Exon 31 is <4 kb in size, although the actual *MUC6* exon 31 is usually >10 kb (often >15 kb). This underscores the extremely incomplete annotation for this VNTR-containing exon. Source: https://genome.ucsc.edu/. Panel **(B)** provides a selective overview of previously reported findings related to *MUC6* exon 31. Panel **(C)** depicts, in cartoon form, the distinction between single-nucleotide variants (SNVs) and variation in a VNTR region. Individual SNVs are by definition small (usually 1 bp), whereas VNTR regions' tandem repeat expansion can lead to a much larger change in the local genomic architecture. For the *MUC6* VNTR region, each repeat unit is ~507 bp in length.

polypeptides help constitute the "gel forming" components of mucus (10).

Another interesting characteristic of the mucin gene family is that the different mucin genes' expression may be subject to orchestrated regulation by inflammation-related transcription factors (11). In cells transcribing the mucin genes clustered together on chromosome 11p15.5, the local chromatin forms three-dimensional structures for coordinated exposure of multiple mucin genes' transcriptional promoter sequences (12). These interactions are complicated (13) and context-specific: In various inflammatory diseases, there is differential regulation of chromosome 11p15.5 mucin genes' expression (14–17).

MUC6 transcripts are expressed at very low levels in human brain (Fig. 2). However, MUC6 protein function may have an impact on the brain. There is increasing appreciation of brain-gut interactions, which may theoretically be affected by the microbiome—that is, intestinal bacteria. Mucins, including *MUC6*, constitute a barrier for intestinal organisms: *MUC6* has been reported to block *Helicobacter pylori* infections (18, 19), and mucin proteins play similar functions in altering risk for other bacterial and viral infections (4, 20–22). The intestinal microbiome may play mechanistic roles in neurodegeneration, including AD (23–25); therefore, a microbiome-modulating gene may affect AD indirectly.

MUC6 may also influence susceptibility to non-GI viral infection, which in theory could be relevant to AD risk. There is a body of literature supporting the hypothesis that AD is caused or exacerbated by viral, particularly herpesvirus, infection (26–29). In a recent study by Readhead et al (30), herpesviruses' genetic imprints were increased in AD brains. Readhead et al also performed a genome-wide association study (GWAS) to assess the associative impact of human genetic variants on vulnerability to brain infection by herpesviruses (30). Remarkably, the human genetic variant "associated with the most viruses. . . (rs71454075) falls within the glycoprotein mucin 6" (30)—and, this virus-associated polymorphism resides in the *MUC6* VNTR region. *MUC6* gene variation was also associated with vulnerability to respiratory syncytial virus infection (31). The association(s) between AD and infectious disease are hypothetical and controversial (see, e.g., [32]). However, the prior published findings suggest that complex processes related to the gut microbiome and/or viral immunity may be modulated by *MUC6*. It is unknown whether the regulation of *MUC6* gene expression affects or is affected by expression of the nearby gene, *AP2A2*.

Adaptor-Related Protein Complex 2 Subunit Alpha 2 (*AP2A2*)

In contrast to *MUC6*, *AP2A2* is transcribed robustly in many human tissues including the brain (Fig. 2). The AP2A2 polypeptide is a component of the adaptor protein complex,

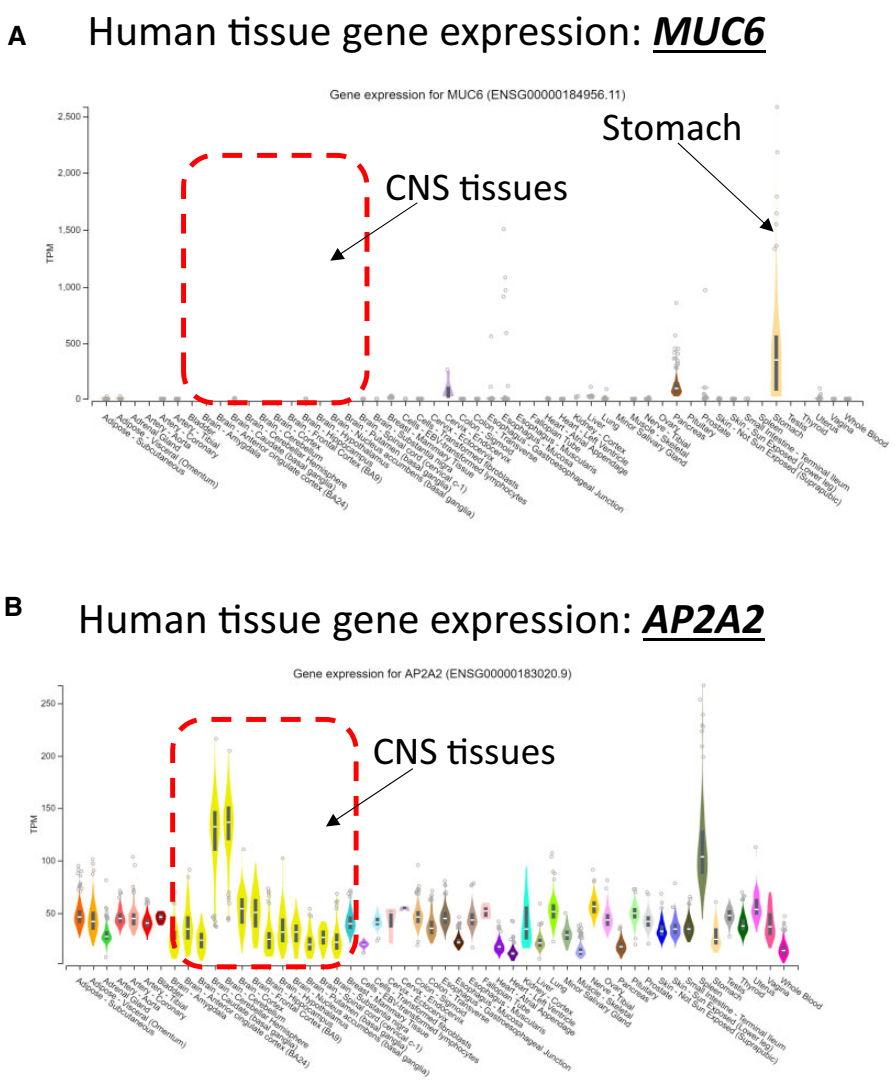

FIGURE 2. Screenshots from the Genotype-Tissue Expression (GTEx) data set depicting human tissue-specific gene expression patterns for *MUC6* **(A)** and *AP2A2* **(B).** Note that *MUC6* is expressed strongly in stomach tissue but not appreciably in the CNS (yellow bars). By contrast, *AP2A2* is expressed at high levels in CNS tissue. The source of the data for these figures was the GTEx Portal, https://gtexportal.org. This figure was adapted from Katsumata et al (1), with permission.

AP-2 (33, 34), a 4-protein multimer that is located on a subset of endocytic vesicles (35, 36). This is an evolutionarily ancient protein complex and the *AP2A2* gene has conserved orthologs in plants, fungi, and invertebrates (37). Functionally, AP-2 participates in assembling the components of early clathrin-coated vesicles (CCVs) at the plasma membrane (33, 35, 38) (Fig. 3). The AP-2 complex interacts directly with lipids, clathrin, "cargos" that are undergoing CCV-mediated endocytosis, and accessory CCV proteins such as PICALM and BIN1 (39–41).

Cargos for CCV-mediated endocytic internalization via AP-2 have been reported to include the amyloid precursor protein (APP), beta-site APP cleaving enzyme 1 (BACE1), Tau, synaptic vesicle proteins, and many others (33, 35, 41–47). It also has been proposed that the apolipoprotein E (APOE) protein and lipoprotein receptors are internalized via CCV endocytosis (43, 48–50). There is an extensive literature documenting the roles of CCV endocytosis and AP-2 in viral infection (51–54). In sum, AP2A2 function may tie in with multiple mechanisms linked to AD pathogenesis.

AP-2 is a stable protein complex comprising 4 domains, which are termed alpha, beta, mu, and sigma (39). *AP2A1* and *AP2A2* are homologous human genes that encode the alpha subunit of the AP-2 complex. Relatively little is known about the molecular neurobiology of *AP2A2 per se*. Publicly accessible databases indicate that the *AP2A2* transcript is expressed in a broad range of CNS cell lineages (data not shown). At the tissue level, both *AP2A1* and *AP2A2* are highly expressed in human cerebellum, but *AP2A2* is expressed at lower levels in the cerebral cortex according to the GTEx Portal

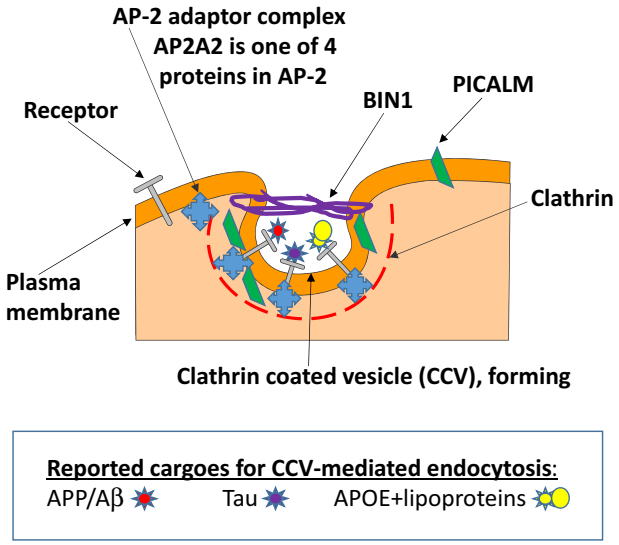

FIGURE 3. Clathrin-coated vesicle (CCV) formation at the plasma membrane involves late-onset Alzheimer's disease (AD) risk-associated gene products including BIN1 and PICALM proteins. The CCV is here shown in early stage of formation at the plasma membrane. Clathrin is depicted with a dashed line. The AP2A2 protein is a component of the AP-2 adapter complex which is located on the plasma membrane and then on CCVs as the vesicle is internalized. Protein cargos that may be related to AD pathogenesis, including Tau, APP, BACE1, APOE, and various viruses have been reported to be processed via CCV-mediated endocytosis. Other (non-CCV) endocytosis mechanisms also occur for some of these proteins. For this cartoon, some relevant structures were minimized for the sake of clarity.

(https://gtexportal.org/ [55]); Figure 4. These findings are intriguing because the cerebellum is relatively resistant to AD-type pathology (56, 57).

We performed immunohistochemical staining for AP2A2 in human brains. In nondemented control brains, AP2A2 immunoreactivity showed a punctate pattern in neurons, consistent with CCV/vesicular structures (1). In AD brains, AP2A2 colocalized with phospho-tau-immunoreactive neurofibrillary tangles (NFTs); by contrast, AP2A2 colocalization with phospho-tau was not seen in progressive supranuclear palsy (PSP)-tufted astrocytes (Fig. 5). Intriguingly, a prior study reported a similar immunostaining pattern using an antibody against CCV-associated protein PICALM—the PICALM protein also was colocalized with phospho-tau in AD NFTs but not with tufted astrocytes in PSP (58).

As to the impact of *AP2A2* genetic variation, prior GWAS have linked *AP2A2* gene variants with AD risk—yet not conclusively (Tables 1 and 2). (By convention, we refer to single nucleotide polymorphisms as SNPs if there is common variability at that locus in human populations and, more generally, as single nucleotide variants [SNVs] if the minor allele is <1% prevalent [59]). A common *AP2A2* SNP (rs10751667) was listed as a "suggestive" AD-linked allele in a large, meta-analytic GWAS: nominal $p \sim 6 \times 10^{-7}$ association with the AD phenotype, not statistically significant after correcting for multiple comparisons (60) (Table 1). In a separate study, the same SNP (rs10751667) was associated with a subtype of mild cognitive impairment (61). Another recent GWAS reported a different *AP2A2* SNP, rs10794342, again with suggestive nominal ($p \sim 4 \times 10^{-6}$) association with AD risk (62) (Table 1).

Larger, meta-analytic AD GWAS have been published in the past few years. We highlight the results of studies by Jansen et al (63) and Kunkle et al (64), for which we obtained summary statistics (Table 2). Again, the associations between *AP2A2* SNPs and the AD phenotype were nominally statistically significant, and with the same risk alleles as in prior reports. A similar finding was reported in the study by Marioni et al (65) (data not shown), but this study sample (the included cases and controls) was highly similar to the later Jansen et al (63) study, for which results are shown in Table 2. More generally, evaluating the abovementioned GWAS independently of each other is not possible via summary statistics because all of these GWAS samples had some degree of overlap.

The 2 marginal AD-associated SNPs (rs10751667 and rs10794342) are in moderate linkage disequilibrium (LD) with each other ($D' = 0.99$, $r^2 = 0.58$ in "all populations" according to ldlink.nci.nih.gov), which means that their genetic variants tend to be co-inherited. In the Nazarian et al (62) and Lambert et al (60) papers, the less common alleles were protective (OR = 0.82 and OR = 0.93, respectively). Variation of both rs10751667 and rs10794342 SNPs is associated with altered expression for *AP2A2*, i.e., both are *AP2A2* expression quantitative trait loci (eQTL), whereas rs10751667 is also an eQTL for *MUC6*. Table 1 shows results from the GTEx Portal website describing meta-analyses of data derived from CNS and non-CNS tissues. As an example of results from brain tissue, according to the GTEx Portal, the rs10794342_T allele is associated with lower *AP2A2* expression in human cerebellum ($p < 1 \times 10^{-7}$ based on the analysis of 209 brain samples). It is an intriguing possibility that a particular polymorphism or haplotype is a risk allele for AD and also could alter expression levels for both *AP2A2* and *MUC6*. The GTEx Portal indicates that rs10794342 and rs10751667 SNPs are also associated with alternative *AP2A2* splicing—thus, are sQTLs for *AP2A2*—in specific human tissues (data not shown), another phenomenon that may merit further study. The SNPs shown in Table 1 are <75 kbp away from the *MUC6* VNTR region, and are in LD with SNPs in the *MUC6* VNTR region (data not shown). Thus, it is possible that the AD-linked *AP2A2* SNPs are a "flag" (proxy feature) for the status of the *MUC6* VNTR.

Additional published data have provided suggestive links between *AP2A2* gene variants and the AD phenotype. A recent GWAS of brain transcriptomic and splicing data, correlating with AD phenotype data, reported that transcript expression variability for both *AP2A2* and the homologous *AP2A1* gene was associated with the AD phenotype (66). A separate study analyzed AD GWAS and transcriptomic data and found that *AP2A2* was a "potential disease-causal" gene (67). Further, in a large, inbred Amish family with genetic risk for AD that was independent of *APOE*, a genomic screen showed that the genetic marker most strongly associated with the AD phenotype (highest multipoint logarithmic of the odds, or

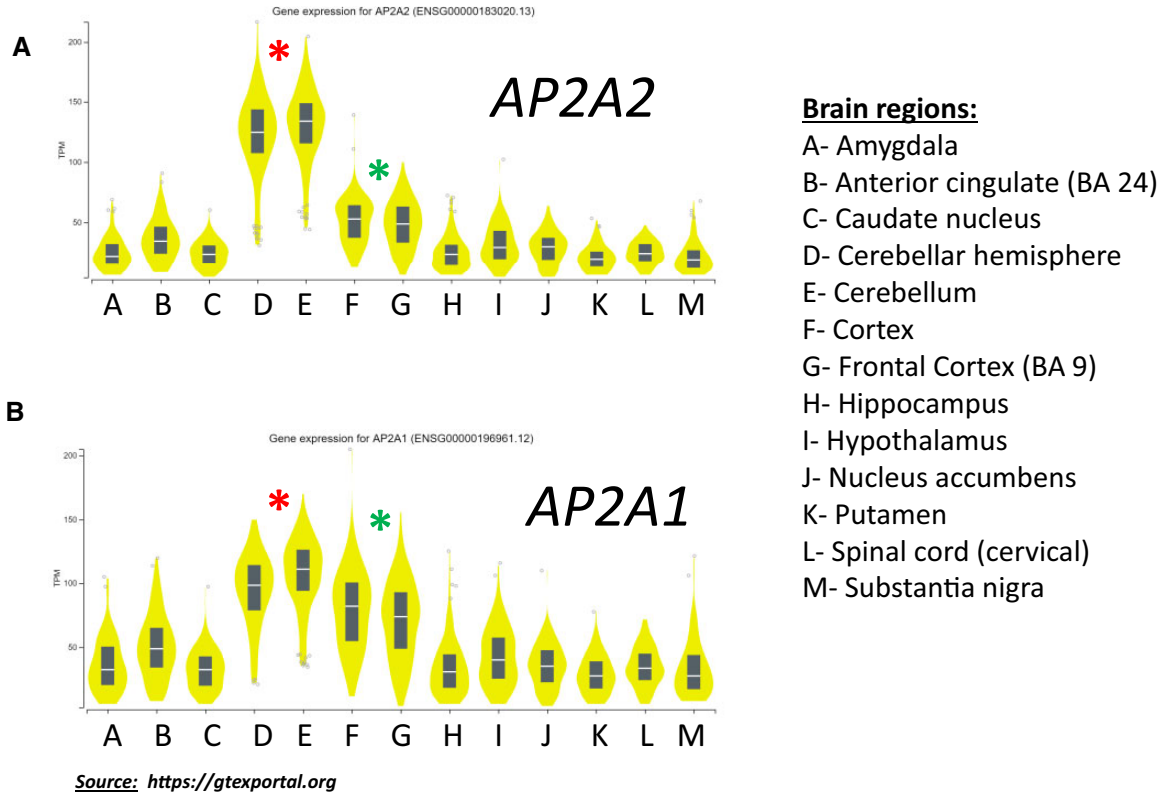

FIGURE 4. Screenshots from the Genotype-Tissue Expression (GTEx) data set depicting human CNS gene expression for the adapter complex (AP-2) alpha subunit-encoding gene transcripts, *AP2A2* **(A)** and *AP2A1* **(B)**. These 2 genes encode proteins homologous in sequence—80.2% identity and 90.3% similar according to https://embnet.vital-it.ch/software/LALIGN_form.html. Both genes' transcripts are expressed at high levels in the cerebellum (red asterisks), but the *AP2A1* (compared to the *AP2A2* transcripts) are expressed at higher levels in the cerebral cortex (green asterisks). Implications of this difference are not known; there have been relatively few studies of *AP2A2* expression regulation in the human brain. BA, Brodmann area. The source for these figures was the GTEx Portal, https://gtexportal.org.

LOD, score) was located <500 kb from *AP2A2* on chromosome 11p (68).

Variants of other genes that encode CCV-related proteins have been repeatedly associated with late-onset AD risk (58, 69–72) (Fig. 3). In a recent large GWAS that incorporated many different study cohorts, *PICALM* and *BIN1* gene variants were the non-*APOE* risk alleles with the lowest p values (73). PICALM and BIN1 proteins have also been implicated in Tau proteinopathy and in APP/Aβ processing (40–42, 58, 74–77). Additional endocytic genes—for example, *SORL1* and *CD2AP*—were also associated with APP processing, autophagy, and clinical dementia (71, 72, 78). Taken together, these results provide theoretical support for the hypothesis that a gene variant which affects AP-2/CCV function could have an impact on AD risk.

The *MUC6* VNTR

The term VNTR describes repetitive end-to-end iterations of genomic sequences ≥6 bp in length (79, 80). The number of repeat units is inherited in the germline and often differs from individual to individual (79, 81). There is a wide variety of VNTR sequences in the human genome (79). We note that the genomic segment we are referring to as a VNTR region could also be referred to as a copy number variant (82, 83), structural variant (84, 85), satellite DNA (86), insertions/deletions (87), simple sequence repeats (88), or tandem repeat expansion, according to different sources. The larger (with more repeats) versions of many different tandemly repeated sequences in the human genome have been associated with neurologic diseases (89), and VNTRs can be transcriptional regulators for nearby genes (90). In the case of the *C9ORF72* tandem repeat expansion, now known to be a strong neurodegenerative disease genetic risk factor (91), the existence of that tandem repeat expansion was signaled by GWAS that identified nearby coinherited SNPs (92–94). Those SNPs are not themselves pathogenetic but they served as proxies to flag the disease-causing tandem repeat locus.

Other investigators have studied the *MUC6* VNTR region (95–101). No other known human tandem repeat sequence is highly similar to the *MUC6* VNTR region. Most human genome VNTR regions contain repeated sequences that are shorter (usually 10–60 bp) than those in the *MUC6* gene (80). By contrast, the tandemly repeated sequences in *MUC6* are

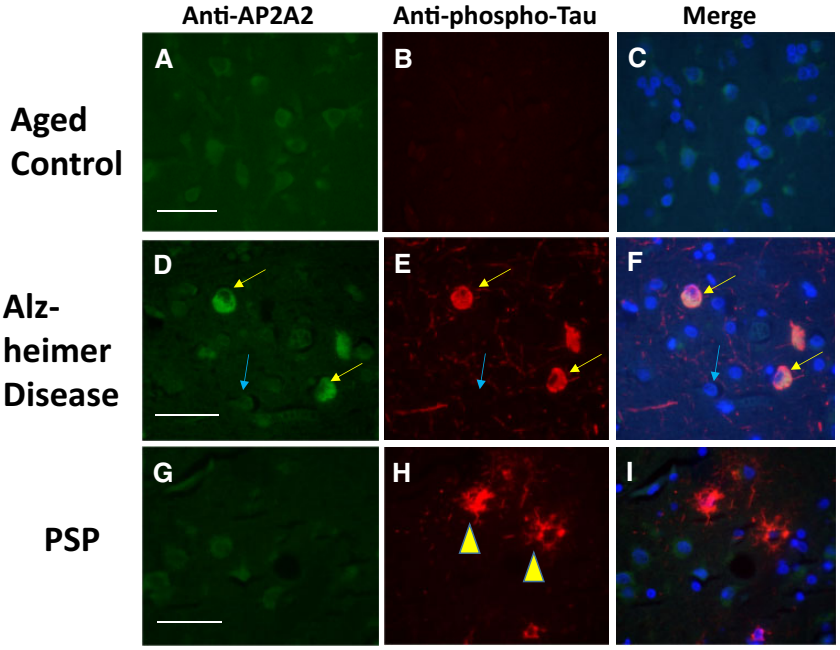

FIGURE 5. Photomicrographs of human brain (aged control **[A–C]**, Alzheimer's disease **[D–F]**, and progressive supranuclear palsy [PSP; **G–I**]) show the colocalization of AP2A2 protein with phospho-tau-immunoreactive neurofibrillary tangles of Alzheimer's disease. Panels **(C)**, **(F)**, and **(I)** have merged staining results combined with blue DAPI counterstain to visualize cell nuclei. AP2A2-immunoreactive cells are seen in aged control brain but are not colocalized with phospho-tau **(A–C)**. In Alzheimer's disease cerebral neocortex, there was a strong tendency for colocalization of an accentuated AP2A2 signal **(D)** that colocalized with phospho-tau **(E)**; see yellow arrows. By contrast, cells that were immunonegative for phospho-tau tangles (blue arrow) had less AP2A2 immunoreactivity. In PSP basal ganglia, the phospho-tau immunoreactivity (yellow arrowheads) in tufted astrocytes did not colocalize with AP2A2 immunoreactivity. The colocalization was quantified and methods published in detail in (1). Thanks to Dr Peter Davies for the anti-phospho-tau (PHF-1) monoclonal antibody. Scale bars = 100 μM.

TABLE 1. *AP2A2* Single Nucleotide Polymorphisms (SNPs) That Showed "Suggestive" Nominal Association With Alzheimer's Disease (AD) Phenotype in Genome-Wide Association Studies (GWAS), Along With Expression Quantitative Trait Locus (eQTL) Status*

AP2A2 SNP With Nominal GWAS Association With the AD Phenotype	Reference	Nominal p Value for Association With AD Phenotype	AD Association Statistically Significant After Correction for Multiple Comparisons?	eQTL for *AP2A2?**	eQTL for *MUC6?**
rs10751667	Lambert et al (60)	$p = 6.3 \times 10^{-7}$	No	Yes; $p = 5.7 \times 10^{-30}$*	Yes; $p = 2.9 \times 10^{-35}$*
rs10794342	Nazarian et al (62)	$p = 4.4 \times 10^{-6}$	No	Yes; $p = 5.9 \times 10^{-40}$*	No*

*eQTL status and p values indicated were according to "Multiple Tissue eQTL Comparison Meta Analysis RE2" at GTEx Portal, https://gtexportal.org/. For details on the methods and numbers of samples involved in the GTEx work, see https://gtexportal.org/home/documentationPage#staticTextAnalysisMethods.

TABLE 2. *AP2A2* Single Nucleotide Polymorphisms (SNPs) From Table 1 in More Recent Alzheimer's Disease GWAS With Larger (but still overlapping) Samples

Reference	Kunkle et al (64)		Jansen et al (63)
	Stage 1	Stage 2	
Sample size, n	63 926*	18 845*	455 258*,†
rs10751667	1.4×10^{-4}‡	2.2×10^{-6}‡	7.5×10^{-3}‡
rs10794342	3.6×10^{-4}‡	9.2×10^{-5}‡	7.2×10^{-3}‡

*Overlapping samples with Lambert et al (60).
†The Jansen et al study included 47 493 Alzheimer's disease "proxy" cases and 328 320 "proxy" controls as determined by reported parental family history.
‡Nominal p values from summary data are shown; these were not statistically significant after correcting for multiple comparisons.

~507 bp each, and these unit repeats are composed of PTS-encoding subunits (79, 95, 101). The number of ~507-bp repeat sequences in the human *MUC6* VNTR region is usually between 15 and 36 repeats (Fig. 1): In ~95% of cases, the *MUC6* VNTR region's size ranges between 7.5 and 18 kb (19, 98). *MUC6* PTS domain-encoding DNA sequences are highly polymorphic (sequences differ between human individuals), causing one group to hypothesize that "the actual amino acid sequence is less important as long as it contains a sufficient number of anchor sites for O-glycans. . ." (95).

Because of the known GI tissue expression of *MUC6*, a major focus in prior work has been on how the *MUC6* tandem repeat polymorphisms are associated with risk for GI

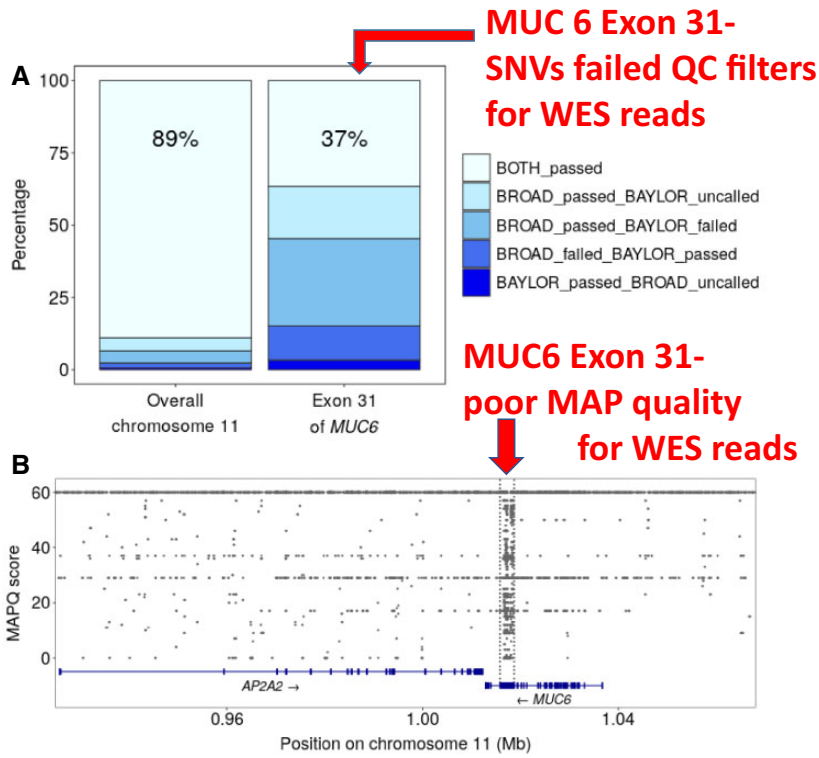

FIGURE 6. The *MUC6* VNTR is a region with poor mapping quality in high-throughput Alzheimer's Disease Sequencing Project (ADSP) whole-exome sequencing (WES) data, resulting in undependable sequencing results. Shown are the percentage of single nucleotide variants (SNVs) that passed quality control (QC) filters in Baylor University (BAYLOR) and Broad Institute (BROAD) analyses **(A)** and mapping quality (MAPQ) score in the *AP2A2*/*MUC6* genomic region **(B)**. The percentages were calculated from variant level-filtering results for the BROAD and BAYLOR pipelines contained in consensus VCF file of ADSP WES data **(A)**. The MAPQ scores were obtained from exome alignment BAM file for subject ID = NA06984 (Utah Residents with Northern and Western European Ancestry) in 1000 Genomes Project Phase 3 **(B)**. Note the poor *MUC6* MAPQ scores in the Exon 31 region. This figure was adapted from Katsumata et al (1), with permission.

diseases—mainly cancer and infection (19, 96, 102, 103). Findings from these previous studies indicated that the short version(s) of the *MUC6* VNTR was the risk alleles, in contrast to what we reported in AD (19, 99). Interestingly, minisatellite polymorphisms outside of exon 31 were also located in and near the *MUC6* gene, and were associated with *MUC6* expression (96, 103). Many other studies have evaluated *MUC6* as a pathologic biomarker in human diseases, but most did not address tandem repeat polymorphism per se.

Unfortunately, the *MUC6* VNTR region is extremely difficult to sequence. This difficulty stems from technical factors, related to the size, number, and interindividual differences of the tandem repeats, as well as the incomplete annotation of the gene (95). There has not been a full-length *MUC6* transcript sequence reported to date (partial cDNA sequences have been described) (7, 10, 97, 100, 104). The technical difficulty in sequencing also applies to DNA haplotypes (79, 95). The annotated GRCh37/GRCh38 reference panel sequences of the *MUC6* VNTR-related haplotypes were imperfectly stitched together (95), and both the reference (Fig. 1) and the alternate contigs (i.e., the standard annotation sequences) are too small, not necessarily representative of any individual, much less a population's repertoire.

In the Alzheimer's Disease Sequencing Project (ADSP) whole-exome sequencing (WES) data set (105, 106), the analytic results reflected the difficulty in sequencing. More specifically, *MUC6* exon 31 produced low mapping quality scores and a commensurate high percentage of SNV calls that failed to pass bioinformatics pipeline quality control filters. For chromosome 11 overall, 89% of SNV calls passed Baylor University (BAYLOR) and Broad Institute (BROAD) filters as defined previously (1), and only 11% failed to pass one or other filter. In sharp contrast, within *MUC6* exon 31, almost two-thirds of SNVs failed both filters (Fig. 6A). Correspondingly, many sequencing reads that were aligned to *MUC6* exon 31 had low mapping quality (Fig. 6B). These data help explain why the *MUC6* VNTR region was largely excluded from many prior studies.

Among other notable features of the mucin genes is their reported potential to vary, within individuals over time, and between human populations. (It is a challenge to determine whether the detected variability signals were due partly to sequencing mistakes.) For example, mucin genes including *MUC6* were on a short list of genes reported to accumulate exomic mutations within individuals during aging (107). In a mucin gene (*MUC2*) that is near *MUC6* on chromosome

11p15.5, the PTS-encoding repeat sequences differed markedly between "healthy Caucasians" and "healthy Asians" (108). Systematic comparisons among diverse groups have not been reported for *MUC6*, although a high degree of *MUC6* VNTR polymorphism has been demonstrated within different populations (19, 99, 109). Variation between diverse populations is a relevant concern because prior studies found that the impacts of specific late-onset AD-risk alleles can vary according to ethnoracial factors (110–114). However, there are abundant scientific and historical reasons to be cautious with research at the nexus of genetics and race, including in dementia research (115).

The Missing/Hidden Heritability Problem in AD Research

A salient question in the field of AD research is: What proportion of risk for developing the AD phenotype is attributable to currently unknown genetic factors? Addressing this involves 2 related questions: (1) What proportion of AD risk is attributable to genetic factors overall? and (2) What proportion of AD risk is explained by known genetic risk factors? There is uncertainty about the answers to all of these questions. Multiple levels of complexity are involved, related to the fields of genomics, neuropathology, epidemiology, and large multicenter studies. Prior reviews and meta-analyses have remarked that the disease-related operationalizations (diagnostic criteria and thresholds/cut-points applied) and also the results of prior AD studies "vary greatly" (116, 117). For all the unanswered questions, there are grounds to hypothesize that a substantial proportion of risk for developing the AD phenotype is due to currently unknown genetic factors—this is the basis for positing a "missing/hidden heritability problem" in AD (118–120).

As to the degree to which AD is explained by genetic variation, there is not a clear consensus from the literature. In a large Swedish twin study, heritability was estimated to explain 58%–79% of AD cases (119). Another group reported that genetic differences explained 53% of phenotypic variance of AD (121). It is also possible that heritability varies across AD subtypes. It was reported that attributable risk due to genetics for episodic memory loss and working memory loss in elderly subjects was approximately 60% and 70%, respectively (122). A study of dysexecutive AD symptoms estimated the heritability of dysexecutive clinical spectrum symptoms (a quantitative measure) to be 68% (123). Different approaches have been used to estimate the impact of specific genetic factors on AD neuropathologic subtypes (124, 125). We emphasize that the heritability of AD is challenging to estimate because data are required from diverse populations (each estimate of heritability may be unique to a particular population, perhaps related to that group's environment [126]), with neuropathologic corroboration, and there are many potential confounders.

The known genetic risk factor with the largest impact on public health is the *APOE* ε4 allele. The AD-associated *APOE* ε4 allele is present in approximately one-fourth of persons, with gene frequency varying among different human populations and greatly enriched in many AD study cohorts (127–129). The *APOE* ε4 allele was initially demonstrated to be associated with the AD phenotype by testing relatively small groups: *APOE* status was compared between 95 AD cases and 139 "unaffected" controls (130).

Extensive GWAS were performed subsequently. Recent AD GWAS sample sizes were large, and imputation approaches are constantly improving (64, 73, 131, 132). These advances enabled the identification of dozens of individual non-*APOE* risk alleles, including common variants with subtle effect sizes and rare alleles with larger effects (60, 64, 73, 133–136). Collectively, and combinatorially, these gene variants may have large impact.

Despite progress in the field, major uncertainties persist. For example, it is not known what proportion of the non-*APOE*-related AD risk is conferred via genetic mechanisms. Determining the scope of this residual genetic risk requires community-based research cohorts because hospital- and clinic-based samples have strong recruitment biases (137–140). In a high-quality community-based study (Rush University Medical Center cohorts), among n = 1017 autopsied subjects, AD pathology was observed in 62.9%, whereas among the same 1017 individuals, only 25.6% had the *APOE* ε4 allele (125). In an analysis of data from the population-based Rotterdam Study, the proportion of incident dementia deemed attributable to the *APOE* ε4 allele was estimated to be 20% (141). These data indicate that a large amount of AD risk is not explained by *APOE*.

Assessing cumulative genetic risk for AD in an individual or a population requires integration of complex genotype data (64, 131, 132). AD polygenic risk scores (PRS) incorporate information about many gene variants to predict an individual's risk of expressing the AD phenotype. PRS explained over 80% of AD risk in selected samples (133, 142–147). Although these studies provided valuable insights into AD genetics, few of these studies had population-representative samples. Some of the prior published papers incorporated data on the variation of thousands of putative risk alleles in the PRS using an unfiltered study design, which may theoretically have included proxies from the *MUC6* region along with many other polymorphisms. Studies that generated a PRS using a predetermined set of well-validated AD-linked SNPs (148–151) found that predicting AD risk using known risk alleles may be useful in various settings (152), but these studies also indicated that there are yet-uncharacterized AD risk factors because a substantial proportion of AD risk was not explained by the known gene variants. According to multiple sources, if one assumes that ∼25% of the population harbors the *APOE* ε4 genotype, then the known common risk variants identified by GWAS explained only 30%–50% of total phenotypic variance of late-onset AD (118, 119, 121, 153–158).

To summarize the current authors' perspective regarding genetic influence on late-onset AD risk: (1) It is currently not known how much of AD risk overall is attributable to genetic factors, but it is often estimated that genetic factors are responsible for ∼50%–80% of AD risk and (2) It also is not known how much of AD risk is attributable to known AD-linked risk alleles, but it may be as low as 30%–50%. With this degree of uncertainty, there are many possibilities, including that there may be currently unknown environmental factors that strongly influence AD risk. Yet as is relevant to the

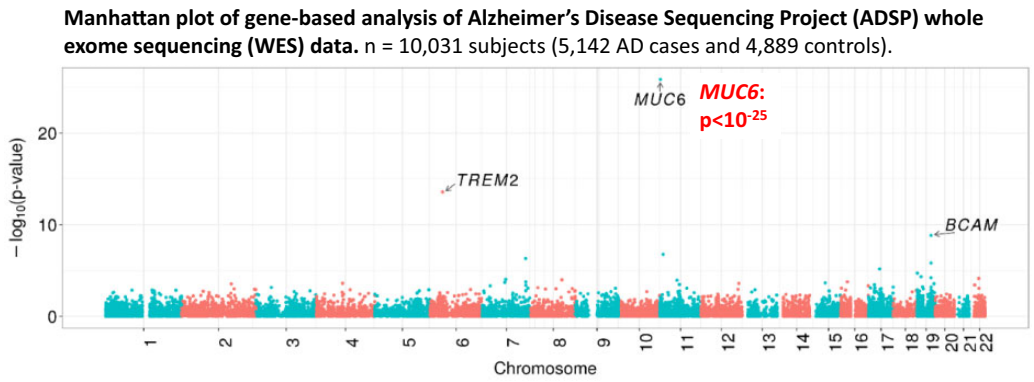

FIGURE 7. Manhattan plot of gene-based analysis of Alzheimer's Disease Sequencing Project (ADSP) whole-exome sequencing (WES) data, showing association with Alzheimer's disease phenotype in that data set. Optimal-unified sequence kernel association test (SKAT-O) association results are depicted for n = 10 031 (5142 AD cases and 4889 controls) subjects. The gene with lowest p value in this genome-wide scan was mucin 6 (*MUC6*), with $p = 1.38 \times 10^{-26}$, and the second lowest p value was triggering receptor expressed on myeloid cells 2 (*TREM2*). Relatively rare alleles in both *TREM2* and basal cell adhesion molecule (*BCAM*), which is located near *APOE* on chromosome 19, were both previously associated with the AD phenotype (106, 176, 178). This figure was adapted from Katsumata et al (1), with permission.

current review article, prior studies also indicate that high-impact AD-associated risk factors may remain to be characterized.

Where Might the Missing Heritability of AD Be Found?

Although increasing GWAS sample sizes and identifying rare disease-associated alleles are potential strategies for filling in the missing/hidden heritability (159–163), the challenges go deeper. Surveying the literature provide indications of the need to develop new approaches for genotype/phenotype correlations. Choosing one example from among many: A relatively easily measured human trait is height. Twin studies indicate that height is highly heritable; ~80% of human height is due to genetic factors (164–166). A recent large GWAS (n = 253 288 included individuals) identified 697 gene variants associated with human height at genome-wide statistical significance (167). These data increased the explanation of the genetic influences on human height from prior studies (168), yet, collectively, these 697 gene variants still explained only one-fifth of the phenotypic (height) variance in the sample (167). Even taking into account "all common variants" only "captured 60% of heritability" (167). Thus, despite large sample sizes and improving imputation methods, some phenotypes may not be fully explained by readily characterized small genetic variants.

One reason for the persistent knowledge gaps is that much of the human genome has remained practically off-limits to existing high-throughput genome-wide sequencing analyses. Related challenges include coping with large structural variants, genomic regions with substantial interindividual variation, and incomplete annotation of the human genome. Further, genomic regions with repetitive sequences were trimmed off by the bioinformatics pipelines used to identify variants in both GWAS and "next-generation sequencing" (NGS) data (169–171).

A recent, directly relevant study by Ebbert et al analyzed ADSP NGS data systematically and reported that there are "many regions with few mappable reads that we call dark by depth, and others that have ambiguous alignment, called camouflaged. . . . The number of genes affected by dark and camouflaged regions was surprisingly high. We identified 36 794 total dark regions across 6054 gene bodies, 3804 of which were protein coding genes. . ." (171).

The discovery of extremely widespread "dark or camouflaged" regions in human genomes has profound implications. This may be particularly true for studies that correlate NGS data with neurologic disease endophenotypes. Even if one hypothesized that genomic regions which can be reliably sequenced are more likely to be disease-driving than the many regions that confound current sequencing technologies, it would still be quite credible that the extensive "dark or camouflaged" regions contain genetic variants which influence human phenotypes. However, for dozens of different human neurologic diseases, a specific tandem repeat-rich genomic region was the established cause of the disease phenotype (172–174). This tendency makes it all the more likely that the missing/hidden heritability for AD could be partly explained by analyzing the widespread, and hitherto poorly characterized, repeat-rich segments of the human genome (83).

The *MUC6/AP2A2* Locus and the AD Endophenotype: Data and Cautionary Considerations

The *MUC6* locus was initially identified to contain a candidate AD-associated polymorphism through analyses of ADSP WES data (n = 5142 AD cases and 4889 control subjects of European ancestry were included) (1). We employed an optimal sequence kernel association test (SKAT-O) (175) to survey genes for SNVs associated with the AD phenotype (1). The results of that gene-based analysis are shown in Figure 7, using data charted in "Manhattan plot" format. Several

aspects of these findings can be appreciated: 1> The reported association between *MUC6* gene variation and the AD phenotype was strong ($p < 10^{-25}$); 2> The other genes associated significantly with the AD phenotype in this analysis agreed with prior published studies (*TREM2* [176, 177] on chromosome 6 and *BCAM* [178], which is located near *APOE* on chromosome 19); and, 3> There was not a haphazard pattern of nominally positive "hits" that would indicate a systematic tendency toward spurious false-positivity.

With regard to the hypothesis that *MUC6* contains an AD-associated polymorphism, there were reasons to be skeptical. Each of the *MUC6* SNVs that was initially found to be associated with the AD phenotype was located within the same unique, poorly annotated, and highly polymorphic VNTR region (1). It was previously demonstrated that the *MUC6* VNTR region is prone to producing false-positive results in WES data because of the extremely high local polymorphism (109, 179). One group suggested that *MUC6* belongs on a "blacklist" of genes that should be viewed with extraordinary skepticism when predicting pathogenicity based on WES results, because of this "excess heterogeneity" (180). The very low nominal p value for the *MUC6*/AD association (Fig. 7) could be interpreted as another indicator that technical bias influenced the results. As stated above, it is possible that the *MUC6* polymorphism could be variable between populations of different ancestry. This needs to be evaluated carefully, because population structure could also introduce false-positive associations. A consideration of these factors makes it seem like a credible hypothesis that *MUC6* genotype/phenotype associations reported by outside authors (e.g., the association between a *MUC6* VNTR gene variant and brain herpesvirus infection [30]) could be false-positives.

Despite these warning signals, we assert that *MUC6* should not be universally or compulsorily ignored. Merely because a gene contains many sequence elements that are highly variable in a population does not mean that all of those genetic variants are equally benign. As specific examples, other genes that are hypothesized to be pathogenetic (e.g., *HLA-DRB1* in multiple sclerosis and AD [181–184] and *FIG4* in amyotrophic lateral sclerosis [185–187]), also were "blacklisted" for extreme heterozygosity, as *MUC6* was (180). It may be argued that *MUC6* deserves added attention because the gene contains a highly polymorphic VNTR region, which has been credibly linked to human diseases (see above). Moreover, thousands of published WES studies did not produce a signal at the *MUC6* locus. The prior literature does, however, indicate that an abundance of caution is merited when studying the association between *MUC6* polymorphism and human disease.

Concerns have also been raised about potential sources of systematic bias in the ADSP WES data set that we used to identify the *MUC6* genotype/AD phenotype association. The *MUC6* SNV calls were derived using bioinformatics pipelines affected by many potential confounders (106, 188). In 2016, the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) posted a topical document titled "Review and Proposed Actions for False-Positive Association Results in ADSP Case-Control Data"; the entire document can be read online (189). Although not mentioning *MUC6* specifically, this posting did refer to the ADSP WES data set version that we had analyzed (phs000572.v6.p4), and stated that "differential effects between cases and controls [led] to strong but spurious associations". Among the notes of caution were: "Different capture protocols were used at the different sequencing centers, and these … could result in biases." And: "Association analysis of case-control genotypes with AD with covariate adjustment for time period of sample sequence processing at the Broad Institute appeared to eliminate most potentially spurious associations." This analysis was complemented by "Sanger sequencing …. to validate heterozygous calls" (189). (The reported Sanger sequencing experiments focused on SNV calls but not on tandem repeat expansion.) Ensuing published work that evaluated AD-linked rare variants in the ADSP WES data set did not mention *MUC6* (171, 176–178).

Consistent with the NIAGADS cautionary posting, we reported experimental evidence that some of the individual *MUC6* SNV calls that had initially focused our interest in the *MUC6* VNTR region were not correct (1). For a small sample of cases, we generated SNV calls via a direct polymerase chain reaction (PCR)-based assay, and the results were different from the SNV calls for the same cases according to the initial ADSP data (1). Based on these findings, if there was a true positive signal of pathogenicity at the *MUC6* locus, it likely was not a scenario where individual disease-causing gene variants were identified.

We conclude that the phenomenon of the *MUC6* gene variants' status being associated with the AD phenotype (Fig. 7) may have been influenced by many factors. As described above, there are reasons to be skeptical: The *MUC6* region's extreme heterozygosity raises the likelihood of a false-positive signal; specific SNV calls in the early ADSP data set were incorrect; and, the initial ADSP analyses were systematically biased. However, the full impact of that bias was not determined, and we underscore the critical distinction between the SNVs which were shown to be problematic and/or mis-called, versus the tandem repeat expansion of *MUC6* that was not addressed by any other workers in the dementia field. The known data are compatible with a testable hypothesis: There was a meaningful underlying biological association between the number of *MUC6* tandem repeats and the AD phenotype. Put another way, the early analyses of WES data may have provided a proxy flag, signaling that *MUC6* tandem repeat expansion is associated with AD risk.

Additional Studies, Preliminary Assessment of Public Health Impact, and Knowledge Gaps

Following the results of ADSP WES data analyses, we used experimental tools tailored specifically for studying the *MUC6* VNTR region: The focus was on the association between AD pathology and *MUC6* tandem repeat expansion, rather than on SNV-type genetic variation. The study sample was the UK-ADC autopsy cohort (190) as described in detail (1). Consistent with prior studies (19, 98), analyzing the autopsied subjects' genomic DNA with PCR revealed a wide range of *MUC6* VNTR sizes. In a Discovery subsample ($n = 119$) of autopsied subjects, the size of the *MUC6* VNTR region was associated with the severity of phospho-tau pathology in

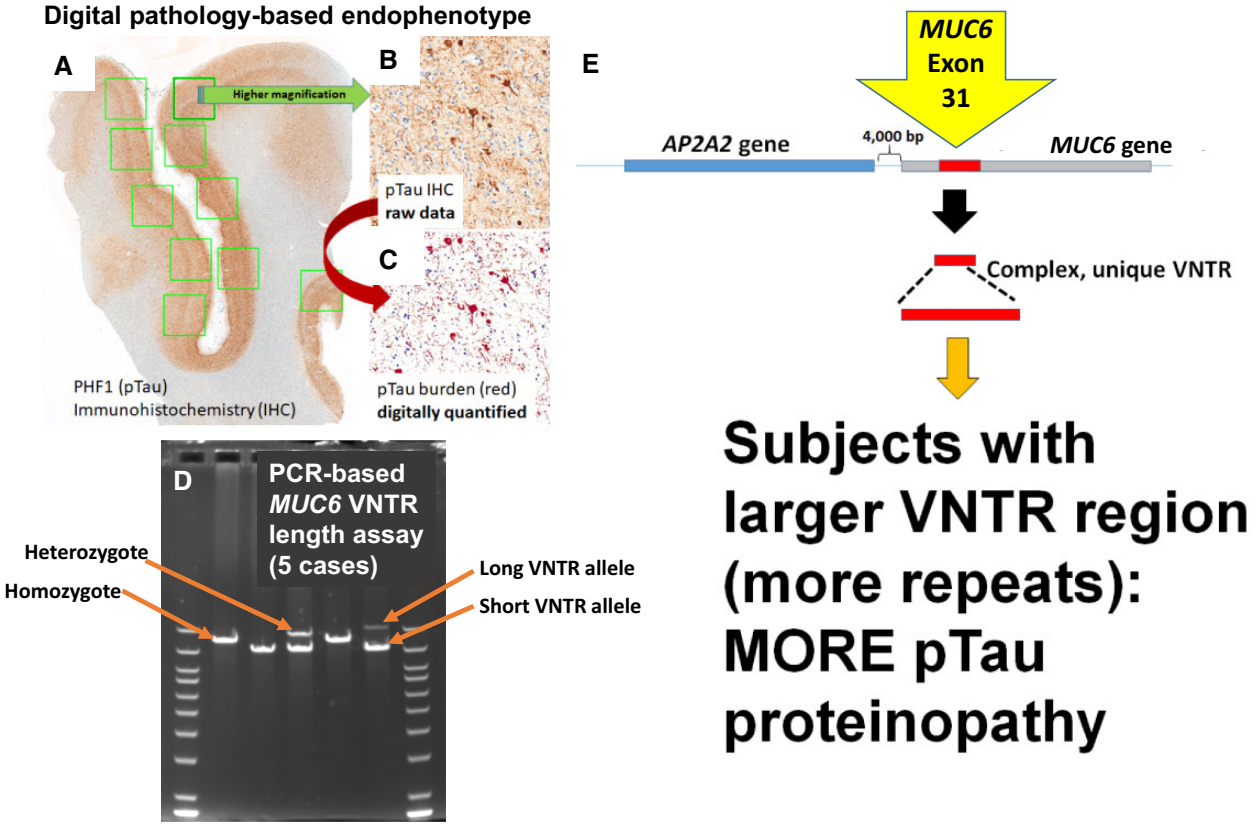

FIGURE 8. Experiments were performed to test the association(s) between Alzheimer's disease phospho-Tau (pTau) pathology, detected with immunohistochemistry **(A–C)**, and the size of the *MUC6* VNTR as measured using a PCR-based assay **(D)**. Digital quantification of pTau pathologic burden was used for correlation with the size of the *MUC6* VNTR polymorphism. Panels **(A–C)** show representative results of pTau immunohistochemistry (IHC) from the temporal neocortex (Brodmann areas 21/22) of a subject with dementia. The actual pTau immunostain is shown in panel **(A)**, and at higher magnification in panel **(B)**. The false-colored data **(C)** conveys the quantified pTau pathology (dark red). Parameters from different brain regions were analyzed to compare the amount of pTau pathology in subjects stratified by the size of the largest detected *MUC6* VNTR region according to PCR **(D)**. PCR results are shown for 5 representative cases; those with one amplicon band on the gel are hypothesized to be homozygous, 2 bands indicate a heterozygote. The slower running (higher on the gel) samples represent the larger alleles. Results are summarized **(E)**. The reason we compared pTau pathology in temporal neocortex in the Replication cohort cases with the largest 33%ile amplicons as measured by PCR (top tertile), to the remaining two-thirds of cases, is that this is the region and threshold in the Discovery cohort with the strongest association between the *MUC6* VNTR region size and pTau pathology (1). This figure was adapted from Katsumata et al (1), with permission.

neocortical brain regions (Fig. 8). For the risk allele, we used as a threshold criterion the largest 33%ile *MUC6* VNTR region size (top tertile) as measured by PCR, because this was the cut-point in the Discovery cohort with the strongest association between the *MUC6* VNTR region size and phospho-tau pathologic severity (1). That main finding was reproduced in a separate Replication cohort (n = 173).

None of our prior published analyses combined the abovementioned Discovery and Replication cohorts from the UK-ADC data set. We here show results of analyses factoring in all of these subjects together. The criterion for operationalizing the *MUC6* risk allele was again that the size of the VNTR region was in the top tertile in that cohort according to the PCR assay. The summary data were analyzed to compare parameters related to AD neuropathologic changes using conventional stage-based AD pathology severity metrics.

Tables 3 and 4 show the results for comparisons of AD neuropathology in the UK-ADC combined cohort, stratifying by *APOE* ε4 allele status and *MUC6* VNTR allele status. Table 3 shows how the *MUC6* VNTR and *APOE* ε4 allele genotypes were correlated with the number of individuals in specific Braak NFT stage (191) groups. There were a total of n = 101 cases with Braak NFT stage VI (which indicates the most severe AD pathology) in this analysis. The assessment of *MUC6* VNTR risk allele status reduced the number of severe AD cases unexplained by a risk allele—the number of *MUC6*−/*APOE*− cases with Braak NFT stage VI was only n = 15. A test using analysis of deviance for a logistic regression model of Braak NFT stage VI, regressed on *MUC6* VNTR status, *APOE* ε4 presence, and their interaction, exhibits high statistical significance (LRT-p = 1.08×10^{-7}). A reciprocal tendency for enrichment of *MUC6*−/*APOE*− cases

TABLE 3. Braak Neurofibrillary Tangle (NFT) Stage Groups in the UK-ADC Autopsy Cohort (1) (n = 292*), Stratifying by *APOE* ε4 Status (*APOE4+/−*) and by Whether or Not the *MUC6* VNTR Region Is in the Top Size Tertile (*MUC6+/−*)

Allele Status†	Total n, Each Group	n, Braak NFT Stage VI Cases	% Braak NFT Stage VI Cases	% Braak NFT Stages 0–III Cases
APOE4−/MUC6−	n=104	n=15	14.4	55.8
APOE4−/MUC6+	n=51	n=22	43.1	31.4
APOE4+/MUC6−	n=71	n=32	45.1	23.9
APOE4+/MUC6+	n=60	n=32	53.3	16.7

*n = 6 had missing Braak NFT stage.
†"+" indicates presence of at least 1 putative risk allele.

TABLE 4. Association Between CERAD Neuritic Amyloid Plaque Density Scores and PCR-Measured *MUC6* VNTR Region Size Stratified by *APOE* ε4 Allele Status in the UK-ADC Cohort (1); n = 292

	Numbers of Subjects by Category		p Value*
	Bottom 2 Tertiles VNTR Size	Top Tertile VNTR Size	
APOE ε4 (−): n = 159			
CERAD none, possible, or probable	n = 72	n = 22	0.0025
CERAD definite (NP density highest)	n = 34	n = 31	
APOE ε4 (+): n = 133			
CERAD none, possible, or probable	n = 26	n = 13	0.12
CERAD definite (NP density highest)	n = 47	n = 47	

*Chi-squared test result, two-tailed test.
CERAD, Consortium to Establish a Registry for Alzheimer's Disease (192).

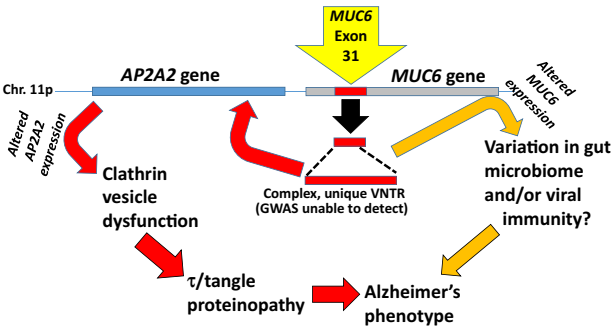

FIGURE 9. Hypotheses related to pathogenetic mechanisms: a variable number of tandem repeat (VNTR) region within the *MUC6* gene on Chr. 11p constitutes a common driver of the Alzheimer's disease (AD) phenotype. The large-sized VNTR (i.e., the risk allele, which is ~30% prevalent in our preliminary studies) is associated with the severity of AD-type pathology, particularly tau neurofibrillary pathology. The *MUC6* VNTR region is immediately downstream of the *AP2A2* gene. The AP2A2 gene product is an "adaptor" protein involved in clathrin-coated vesicle (CCV) formation at the plasma membrane (33, 35, 38). We found that *AP2A2* expression is dysregulated in persons with the large-sized VNTR region (1). There are also lines of evidence connecting *MUC6* with the gut microbiome and viral immunity as discussed in the text of the present review; these inflammatory influences may contribute to altered risk of AD.

can be observed among subjects with relatively low neurofibrillary pathology, Braak NFT stages 0–III. Table 4 shows the results for neuritic amyloid plaques, graded according to the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuritic plaque density scores (192). Among the subjects who lacked the *APOE* ε4 allele, carrying the *MUC6* VNTR risk allele correlated with having highest grade density of neuritic amyloid plaques. The same general trend was present among those with the *APOE* ε4 allele. This is the first demonstration of amyloid plaque-associated pathology correlated with the *MUC6* VNTR genotype.

Published data (1) and a literature review provide support for hypotheses that link the *MUC6* VNTR region polymorphism with AD pathogenesis (Fig. 9). The expanded version of the *MUC6* VNTR region was associated with decreased *AP2A2* expression (1). *MUC6* VNTR polymorphism thus may affect CCV function, which is related to autophagy and functions in processing neurodegenerative disease-associated proteins such as APP and Tau (41, 42, 75). Alternatively, or in parallel, the combination of impacts on *MUC6* and *AP2A2* could influence the gut microbiome and/or viral immunity, therefore affecting pathogenetic cascades relevant to AD. The long-term effects of the variation of the *MUC6* VNTR region may cause or accentuate pathologies, including tau neurofibrillary degeneration and neuritic amyloid plaques, in aged persons' brains.

It is premature to make generalizations about the influence of the *MUC6* VNTR expansion on public health. For now, we can only analyze the relatively small UK-ADC autopsy cohort sample. We performed preliminary calculations of the AD odds ratio (OR), using the combined sample from the UK-ADC autopsy cohort (n = 292); see Figure 10. The *MUC6* VNTR risk allele is ~35% prevalent (top tertile) in this

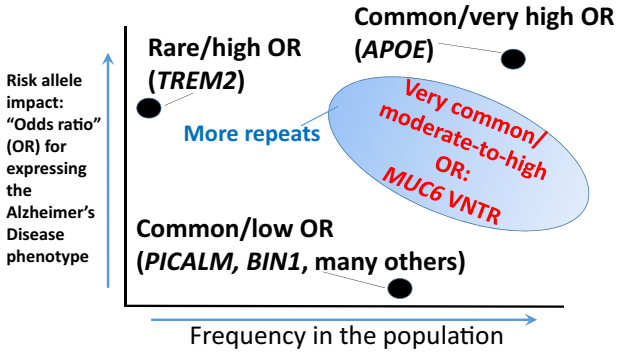

FIGURE 10. Preliminary estimations of public health impact of the *MUC6* VNTR polymorphism: risk allele prevalence and Alzheimer's disease (AD) "odds ratio" (OR). According to our preliminary data in the University of Kentucky AD Center cohort, the AD OR for the top-tertile risk allele is 2.10 (95% CI 1.07–4.18) using Tau pathology in the temporal neocortex (Brodmann areas 21/22) as the endophenotype, compared to *APOE* ε4 OR = 2.87 (95% CI 1.55–5.58). In prior work, other non-*APOE* common late-onset AD-risk alleles' OR were <1.5 (60). The UK-ADC cohort is a relatively small sample that lacks socioeconomic or ethnoracial diversity. Additional studies are required to better understand the public health impact of the *MUC6* VNTR polymorphism.

cohort, and the *MUC6* VNTR risk allele OR for AD is 2.10 (95% CI 1.07–4.18) using the severity of temporal neocortex NFT pathology as the endophenotype, compared to *APOE* ε4 OR = 2.87 (95% CI 1.55–5.58) for the same endophenotype and in the same cohort.

Among tested subjects in the UK-ADC sample, persons with the *APOE* ε4 allele were relatively likely to also have the *MUC6* VNTR risk allele (Table 3): 32.8% of *APOE* ε4(−) subjects had the *MUC6* VNTR risk allele, whereas 45.8% of the *APOE* ε4(+) subjects had the *MUC6* VNTR risk allele. Although this was only a marginal trend ($p < 0.03$) in this small cohort, this phenomenon is notable. Primarily, to evaluate the impact of the *MUC6* VNTR risk allele appropriately, we factored each subject's *APOE* ε4 allele status into our analytic model (1), so as not to overestimate the associative impact of the *MUC6* VNTR risk allele.

The UK-ADC sample's enrichment of subjects with both the *MUC6* VNTR risk allele and (in the same persons) the *APOE* ε4 allele may have other implications. As noted previously, AD-related study samples are often influenced by recruitment bias to include a disproportionately high number of individuals at genetic risk for AD-type dementia. (The allele frequency for the *APOE* ε4 allele was >45% in this dementia-enriched UK-ADC convenience sample, although the population frequency is ~25%.) If the *MUC6* VNTR risk allele was indeed influencing AD risk, one would hypothesize that there should be a corresponding tendency to recruit high-risk subjects with both the large *MUC6* VNTR and the *APOE* ε4 allele; this bias may not be restricted to the UK-ADC cohort. If this were true, then, an additional hypothesis would follow: Just as an estimate of the impact of the *MUC6* VNTR must be adjusted to account for the presence of the *APOE* ε4 allele in carriers of both risk alleles, so should an estimate of the impact of the *APOE* ε4 allele be adjusted to correct for the presence of the *MUC6* VNTR risk allele. In other words, a portion of AD risk which hitherto has been attributed to *APOE*, may actually have been due to the *MUC6* VNTR allele, because dementia research study subjects may tend to harbor both risk alleles simultaneously due to recruitment bias. Further, persons with the *APOE* ε4 allele but with a risk mitigating (short) *MUC6* VNTR allele may theoretically be less likely to be represented in dementia research cohorts.

Although the data gathered related to the *MUC6* VNTR and AD risk are suggestive so far, there are also reasons to be cautious, in addition to those mentioned above. The sample that we assessed was relatively small by genomics study standards. These findings require further corroboration. The UK-ADC autopsy cohort is a relatively homogenous demographic sample, so there is a need for assessment of more diverse and population-representative cohorts. There are other key knowledge gaps that may help guide future studies. For example, we do not know the actual *MUC6* VNTR sequences for any of the cases we have analyzed. To characterize the *MUC6* VNTR polymorphism, a PCR-based assay was developed which indicated the approximate size of the *MUC6* VNTR (Fig. 8D), because we lacked access to a quantitative tandem repeat assay such as is available for *C9ORF72* repeat number characterization (193, 194). An improved *MUC6* VNTR assay or sequencing platform is required for both research and possible future clinical purposes. At this locus and others, there remains much to be learned about the "dark and camouflaged" (171) regions of the human genome.

REFERENCES

1. Katsumata Y, Fardo DW, Bachstetter AD, et al. Alzheimer disease pathology-associated polymorphism in a complex variable number of tandem repeat region within the MUC6 gene, near the AP2A2 gene. J Neuropathol Exp Neurol 2019;79:3–21
2. Ambrosini A, Paul S, Hu S, et al. Human subtelomeric duplicon structure and organization. Genome Biol 2007;8:R151
3. Rousseau K, Byrne C, Griesinger G, et al. Allelic association and recombination hotspots in the mucin gene (MUC) complex on chromosome 11p15.5. Ann Human Genet 2007;71:561–9
4. Corfield AP. Mucins: A biologically relevant glycan barrier in mucosal protection. Biochim Biophys Acta 2015;1850:236–52
5. Ho SB, Roberton AM, Shekels LL, et al. Expression cloning of gastric mucin complementary DNA and localization of mucin gene expression. Gastroenterology 1995;109:735–47
6. Gipson IK, Ho SB, Spurr-Michaud SJ, et al. Mucin genes expressed by human female reproductive tract epithelia. Biol Reprod 1997;56:999–1011
7. Seregni E, Botti C, Massaron S, et al. Structure, function and gene expression of epithelial mucins. Tumori 1997;83:625–32
8. Bartman AE, Buisine MP, Aubert JP, et al. The MUC6 secretory mucin gene is expressed in a wide variety of epithelial tissues. J Pathol 1998;186:398–405
9. Toribara NW, Roberton AM, Ho SB, et al. Human gastric mucin. Identification of a unique species by expression cloning. J Biol Chem 1993;268:5879–85
10. Moniaux N, Escande F, Porchet N, et al. Structural organization and classification of the human mucin genes. Front Biosci 2001;6:D1192–206
11. Van Seuningen I, Pigny P, Perrais M, et al. Transcriptional regulation of the 11p15 mucin genes. Towards new biological tools in human therapy, in inflammatory diseases and cancer? Front Biosci 2001;6:D1216–34
12. Gosalia N, Leir SH, Harris A. Coordinate regulation of the gel-forming mucin genes at chromosome 11p15.5. J Biol Chem 2013;288:6717–25

13. Voynow JA, Gendler SJ, Rose MC. Regulation of mucin genes in chronic inflammatory airway diseases. Am J Respir Cell Mol Biol 2006;34:661–5
14. Niv Y. *Helicobacter pylori* and gastric mucin expression: A systematic review and meta-analysis. World J Gastroenterol 2015;21:9430–6
15. Jankovic Velickovic L, Katic V, Hattori T, et al. Differences in the expression of mucins in various forms of cystitis glandularis. Pathol Res Pract 2007;203:653–8
16. Buisine MP, Desreumaux P, Leteurtre E, et al. Mucin gene expression in intestinal epithelial cells in Crohn's disease. Gut 2001;49:544–51
17. de Bolos C, Real FX, Lopez-Ferrer A. Regulation of mucin and glycoconjugate expression: From normal epithelium to gastric tumors. Front Biosci 2001;6:D1256–63
18. Kawakubo M, Ito Y, Okimura Y, et al. Natural antibiotic function of a human gastric mucin against *Helicobacter pylori* infection. Science 2004;305:1003–6
19. Nguyen TV, Janssen M Jr, Gritters P, et al. Short mucin 6 alleles are associated with *H. pylori* infection. World J Gastroenterol 2006;12: 6021–5
20. McGuckin MA, Linden SK, Sutton P, et al. Mucin dynamics and enteric pathogens. Nat Rev Microbiol 2011;9:265–78
21. Cone RA. Barrier properties of mucus. Adv Drug Deliv Rev 2009;61: 75–85
22. Linden SK, Sutton P, Karlsson NG, et al. Mucins in the mucosal barrier to infection. Mucosal Immunol 2008;1:183–97
23. Shen L, Liu L, Ji HF. Alzheimer's disease histological and behavioral manifestations in transgenic mice correlate with specific gut microbiome state. J Alzheimers Dis 2017;56:385–90
24. Kohler CA, Maes M, Slyepchenko A, et al. The gut-brain axis, including the microbiome, leaky gut and bacterial translocation: Mechanisms and pathophysiological role in Alzheimer's disease. Curr Pharm Des 2016;22:6152–66
25. Bhattacharjee S, Lukiw WJ. Alzheimer's disease and the microbiome. Front Cell Neurosci 2013;7:153
26. Ball MJ. The essential lesion of Alzheimer disease: A surprise in retrospect. J Alzheimers Dis 2006;9:29–33
27. Haas JG, Lathe R. Microbes and Alzheimer's disease: New findings call for a paradigm change. Trends Neurosci 2018;41:570–3
28. Hill JM, Zhao Y, Clement C, et al. HSV-1 infection of human brain cells induces miRNA-146a and Alzheimer-type inflammatory signaling. Neuroreport 2009;20:1500–5
29. Piacentini R, De Chiara G, Li Puma DD, et al. HSV-1 and Alzheimer's disease: More than a hypothesis. Front Pharmacol 2014;5:97
30. Readhead B, Haure-Mirande JV, Funk CC, et al. Multiscale analysis of independent Alzheimer's cohorts finds disruption of molecular, genetic, and clinical networks by human herpesvirus. Neuron 2018;99:64–82 e7
31. Salas A, Pardo-Seco J, Cebey-Lopez M, et al. Whole exome sequencing reveals new candidate genes in host genomic susceptibility to respiratory syncytial virus disease. Sci Rep 2017;7:15888
32. Allnutt MA, Johnson K, Bennett DA, et al. Human herpesvirus 6 detection in Alzheimer's disease cases and controls across multiple cohorts. Neuron 2020;105:1027–35 e2
33. McMahon HT, Boucrot E. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. Nat Rev Mol Cell Biol 2011;12:517–33
34. Boehm M, Bonifacino JS. Adaptins: The final recount. Mol Biol Cell 2001;12:2907–20
35. Rodemer C, Haucke V. Clathrin/AP-2-dependent endocytosis: A novel playground for the pharmacological toolbox? Handb Exp Pharmacol 2008;186:105–22
36. Ohno H. Physiological roles of clathrin adaptor AP complexes: Lessons from mutant animals. J Biochem 2006;139:943–8
37. Dergai M, Iershov A, Novokhatska O, et al. Evolutionary changes on the way to clathrin-mediated endocytosis in animals. Genome Biol Evol 2016;8:588–606
38. Kratzke M, Candiello E, Schmidt B, et al. AP-1/sigma1B-dependent SV protein recycling is regulated in early endosomes and is coupled to AP-2 endocytosis. Mol Neurobiol 2015;52:142–61
39. Hirst J, Robinson MS. Clathrin and adaptors. Biochim Biophys Acta 1998;1404:173–93
40. Malki I, Cantrelle FX, Sottejeau Y, et al. Regulation of the interaction between the neuronal BIN1 isoform 1 and tau proteins – role of the SH3 domain. FEBS J 2017;284:3218–29
41. Tian Y, Chang JC, Fan EY, et al. Adaptor complex AP2/PICALM, through interaction with LC3, targets Alzheimer's APP-CTF for terminal degradation via autophagy. Proc Natl Acad Sci U S A 2013;110: 17071–6
42. Azarnia Tehran D, Kuijpers M, Haucke V. Presynaptic endocytic factors in autophagy and neurodegeneration. Curr Opin Neurobiol 2018; 48:153–9
43. Van Acker ZP, Bretou M, Annaert W. Endo-lysosomal dysregulations and late-onset Alzheimer's disease: Impact of genetic risk factors. Mol Neurodegener 2019;14:20
44. Azarnia Tehran D, Lopez-Hernandez T, Maritzen T. Endocytic adaptor proteins in health and disease: Lessons from model organisms and human mutations. Cells 2019;8:1–52
45. Poulsen ET, Iannuzzi F, Rasmussen HF, et al. An aberrant phosphorylation of amyloid precursor protein tyrosine regulates its trafficking and the binding to the clathrin endocytic complex in neural stem cells of Alzheimer's disease patients. Front Mol Neurosci 2017;10:59
46. Marcello E, Saraceno C, Musardo S, et al. Endocytosis of synaptic ADAM10 in neuronal plasticity and Alzheimer's disease. J Clin Invest 2013;123:2523–38
47. Prabhu Y, Burgos PV, Schindler C, et al. Adaptor protein 2-mediated endocytosis of the beta-secretase BACE1 is dispensable for amyloid precursor protein processing. Mol Biol Cell 2012;23:2339–51
48. Holtzman DM, Herz J, Bu G. Apolipoprotein E and apolipoprotein E receptors: Normal biology and roles in Alzheimer disease. Cold Spring Harb Perspect Med 2012;2:a006312
49. Boucrot E, Saffarian S, Zhang R, et al. Roles of AP-2 in clathrin-mediated endocytosis. PLoS One 2010;5:e10597
50. Garuti R, Jones C, Li WP, et al. The modular adaptor protein autosomal recessive hypercholesterolemia (ARH) promotes low density lipoprotein receptor clustering into clathrin-coated pits. J Biol Chem 2005;280: 40996–1004
51. Beacham GM, Partlow EA, Hollopeter G. Conformational regulation of AP1 and AP2 clathrin adaptor complexes. Traffic 2019;20:741–51
52. Agrelli A, de Moura RR, Crovella S, et al. ZIKA virus entry mechanisms in human cells. Infect Genet Evol 2019;69:22–9
53. Mercer J, Schelhaas M, Helenius A. Virus entry by endocytosis. Annu Rev Biochem 2010;79:803–33
54. Byland R, Vance PJ, Hoxie JA, et al. A conserved dileucine motif mediates clathrin and AP-2-dependent endocytosis of the HIV-1 envelope protein. Mol Biol Cell 2007;18:414–25
55. Consortium GT. The genotype-tissue expression (GTEx) project. Nat Genet 2013;45:580–5
56. Braak H, Braak E. Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol Aging 1995;16:271–8, Discussion 8–84
57. Thal DR, Rub U, Orantes M, et al. Phases of A beta-deposition in the human brain and its relevance for the development of AD. Neurology 2002;58:1791–800
58. Ando K, Tomimura K, Sazdovitch V, et al. Level of PICALM, a key component of clathrin-mediated endocytosis, is correlated with levels of phosphotau and autophagy-related proteins and is associated with tau inclusions in AD, PSP and Pick disease. Neurobiol Dis 2016;94:32–43
59. Haraksingh RR, Snyder MP. Impacts of variation in the human genome on gene regulation. J Mol Biol 2013;425:3970–7
60. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013;45:1452–8
61. Espinosa A, Hernandez-Olasagarre B, Moreno-Grau S, et al. Exploring genetic associations of Alzheimer's disease loci with mild cognitive impairment neurocognitive endophenotypes. Front Aging Neurosci 2018; 10:340
62. Nazarian A, Yashin AI, Kulminski AM. Genome-wide analysis of genetic predisposition to Alzheimer's disease and related sex disparities. Alzheimers Res Ther 2019;11:5
63. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet 2019;51:404–13

64. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nat Genet 2019; 51:414–30
65. Marioni RE, Harris SE, Zhang Q, et al. GWAS on family history of Alzheimer's disease. Transl Psychiatry 2018;8:99
66. Raj T, Li YI, Wong G, et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. Nat Genet 2018;50:1584–92
67. Hao S, Wang R, Zhang Y, et al. Prediction of Alzheimer's disease-associated genes by integration of GWAS summary data and expression data. Front Genet 2018;9:653
68. Ashley-Koch AE, Shao Y, Rimmler JB, et al. An autosomal genomic screen for dementia in an extended Amish family. Neurosci Lett 2005; 379:199–204
69. Gao S, Casey AE, Sargeant TJ, et al. Genetic variation within endolysosomal system is associated with late-onset Alzheimer's disease. Brain 2018;141:2711–20
70. Morgan K. The three new pathways leading to Alzheimer's disease. Neuropathol Appl Neurobiol 2011;37:353–7
71. Kimura N, Yanagisawa K. Traffic jam hypothesis: Relationship between endocytic dysfunction and Alzheimer's disease. Neurochem Int 2018;119:35–41
72. Small SA, Simoes-Spassov S, Mayeux R, et al. Endosomal traffic jams represent a pathogenic hub and therapeutic target in Alzheimer's disease. Trends Neurosci 2017;40:592–602
73. Andrews SJ, Fulton-Howard B, Goate A. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. Lancet Neurol 2020; doi: 10.1016/S1474-4422(19)30435-1
74. Holler CJ, Davis PR, Beckett TL, et al. Bridging integrator 1 (BIN1) protein expression increases in the Alzheimer's disease brain and correlates with neurofibrillary tangle pathology. J Alzheimers Dis 2014;42: 1221–7
75. Poulsen ET, Larsen A, Zollo A, et al. New insights to clathrin and adaptor protein 2 for the design and development of therapeutic strategies. Int J Mol Sci 2015;16:29446–53
76. Baig S, Joseph SA, Tayler H, et al. Distribution and expression of picalm in Alzheimer disease. J Neuropathol Exp Neurol 2010;69: 1071–7
77. Heckmann BL, Teubner BJW, Tummers B, et al. LC3-associated endocytosis facilitates beta-amyloid clearance and mitigates neurodegeneration in murine Alzheimer's disease. Cell 2019;178:536–51.e14
78. Juul Rasmussen I, Tybjærg-Hansen A, Rasmussen KL, et al. Blood-brain barrier transcytosis genes, risk of dementia and stroke: A prospective cohort study of 74,754 individuals. Eur J Epidemiol 2019;34:579–90
79. Bakhtiari M, Shleizer-Burko S, Gymrek M, et al. Targeted genotyping of variable number tandem repeats with adVNTR. Genome Res 2018; 28:1709–19
80. Naslund K, Saetre P, von Salome J, et al. Genome-wide prediction of human VNTRs. Genomics 2005;85:24–35
81. Brookes KJ. The VNTR in complex disorders: The forgotten polymorphisms? A functional way forward? Genomics 2013;101:273–81
82. Lauer S, Gresham D. An evolving view of copy number variants. Curr Genet 2019;65:1287–95
83. Szigeti K. New genome-wide methods for elucidation of candidate copy number variations (CNVs) contributing to Alzheimer's disease heritability. Methods Mol Biol 2016;1303:315–26
84. Mahmoud M, Gobet N, Cruz-Davalos DI, et al. Structural variant calling: The long and the short of it. Genome Biol 2019;20:246
85. Ho SS, Urban AE, Mills RE. Structural variation in the sequencing era. Nat Rev Genet 2020;21:171–89
86. Lower SS, McGurk MP, Clark AG, et al. Satellite DNA evolution: Old ideas, new approaches. Curr Opin Genet Dev 2018;49:70–8
87. Karakoc E, Alkan C, O'Roak BJ, et al. Detection of structural variants and indels within exome data. Nat Methods 2012;9:176–8
88. Roses AD, Akkari PA, Chiba-Falek O, et al. Structural variants can be more informative for disease diagnostics, prognostics and translation than current SNP mapping and exon sequencing. Expert Opin Drug Metab Toxicol 2016;12:135–47
89. Ciesiolka A, Jazurek M, Drazkowska K, et al. Structural characteristics of simple RNA repeats associated with disease and their deleterious protein interactions. Front Cell Neurosci 2017;11:97
90. Nakamura Y, Koyama K, Matsushima M. VNTR (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators. J Hum Genet 1998;43:149–52
91. Marogianni C, Rikos D, Provatas A, et al. The role of C9orf72 in neurodegenerative disorders: A systematic review, an updated meta-analysis, and the creation of an online database. Neurobiol Aging 2019;84:238 e25–e34
92. Laaksovirta H, Peuralinna T, Schymick JC, et al. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: A genome-wide association study. Lancet Neurol 2010;9:978–85
93. Shatunov A, Mok K, Newhouse S, et al. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: A genome-wide association study. Lancet Neurol 2010;9:986–94
94. van Es MA, Veldink JH, Saris CG, et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat Genet 2009;41:1083–7
95. Svensson F, Lang T, Johansson MEV, et al. The central exons of the human MUC2 and MUC6 mucins are highly repetitive and variable in sequence between individuals. Sci Rep 2018;8:17503
96. Ahn MH, Bae KB, Kwon JA, et al. Association of MUC6-minisatellite variants with susceptibility to rectal carcinoma. Mol Biol Rep 2013;40: 303–8
97. Parry S, Sutton-Smith M, Heal P, et al. Evaluation of MUC6 mucin tandem repeats. Biochim Biophys Acta 2005;1722:77–83
98. Vinall LE, Hill AS, Pigny P, et al. Variable number tandem repeat polymorphism of the mucin genes located in the complex on 11p15.5. Hum Genet 1998;102:357–66
99. Garcia E, Carvalho F, Amorim A, et al. MUC6 gene polymorphism in healthy individuals and in gastric cancer patients from northern Portugal. Cancer Epidemiol Biomarkers Prev 1997;6:1071–4
100. Bobek LA, Tsai H, Biesbrock AR, et al. Molecular cloning, sequence, and specificity of expression of the gene encoding the low molecular weight human salivary mucin (MUC7). J Biol Chem 1993;268:20563–9
101. Toribara NW, Ho SB, Gum E, et al. The carboxyl-terminal sequence of the human secretory mucin, MUC6. Analysis Of the primary amino acid sequence. J Biol Chem 1997;272:16398–403
102. Yamazaki K, Tajima Y, Makino R, et al. Tumor differentiation phenotype in gastric differentiated-type tumors and its relation to tumor invasion and genetic alterations. World J Gastroenterol 2006;12:3803–9
103. Kwon JA, Lee SY, Ahn EK, et al. Short rare MUC6 minisatellites-5 alleles influence susceptibility to gastric carcinoma by regulating gene. Hum Mutat 2010;31:942–9
104. Troxler RF, Iontcheva I, Oppenheim FG, et al. Molecular characterization of a major high molecular weight mucin from human sublingual gland. Glycobiology 1997;7:965–73
105. Crane PK, Foroud T, Montine TJ, et al. Alzheimer's Disease Sequencing Project discovery and replication criteria for cases and controls: Data from a community-based prospective cohort study with autopsy follow-up. Alzheimers Dement 2017;13:1410–3
106. Naj AC, Lin H, Vardarajan BN, et al. Quality control and integration of genotypes from two calling pipelines for whole genome sequence data in the Alzheimer's disease sequencing project. Genomics 2019;111: 808–18
107. Bavarva JH, Tae H, McIver L, et al. The dynamic exome: Acquired variants as individuals age. Aging 2014;6:511–21
108. Toribara NW, Gum JR Jr, Culhane PJ, et al. MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorphism. J Clin Invest 1991;88:1005–13
109. Kwak SH, Chae J, Choi S, et al. Findings of a 1303 Korean whole-exome sequencing study. Exp Mol Med 2017;49:e356
110. Huang P, Hsieh SW, Chang YH, et al. Differences in the frequency of Alzheimer's disease-associated genomic variations in populations of different races. Geriatr Gerontol Int 2017;17:2184–93
111. Vardarajan BN, Barral S, Jaworski J, et al. Whole genome sequencing of Caribbean Hispanic families with late-onset Alzheimer's disease. Ann Clin Transl Neurol 2018;5:406–17
112. Reitz C, Mayeux R. Genetics of Alzheimer's disease in Caribbean Hispanic and African American populations. Biol Psychiatry 2014;75: 534–41
113. Miyashita A, Koike A, Jun G, et al. SORL1 is genetically associated with late-onset Alzheimer's disease in Japanese, Koreans and Caucasians. PLoS One 2013;8:e58618

114. Logue MW, Lancour D, Farrell J, et al. Targeted sequencing of Alzheimer disease genes in African Americans implicates novel risk variants. Front Neurosci 2018;12:592
115. Ighodaro ET, Nelson PT, Kukull WA, et al. Challenges and considerations related to studying dementia in Blacks/African Americans. J Alzheimers Dis 2017;60:1–10
116. Stocker H, Mollers T, Perna L, et al. The genetic risk of Alzheimer's disease beyond APOE epsilon4: Systematic review of Alzheimer's genetic risk scores. Transl Psychiatry 2018;8:166
117. Tasaki S, Gaiteri C, Mostafavi S, et al. The molecular and neuropathological consequences of genetic risk for Alzheimer's dementia. Front Neurosci 2018;12:699
118. Ridge PG, Mukherjee S, Crane PK, et al. Alzheimer's disease: Analyzing the missing heritability. PLoS One 2013;8:e79771
119. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry 2006;63:168–74
120. Sun J, Song F, Wang J, et al. Hidden risk genes with high-order intragenic epistasis in Alzheimer's disease. J Alzheimers Dis 2014;41:1039–56
121. Ridge PG, Hoyt KB, Boehme K, et al. Assessment of the genetic variance of late-onset Alzheimer's disease. Neurobiol Aging 2016;41:200 e13–20
122. Wilson RS, Barral S, Lee JH, et al. Heritability of different forms of memory in the late onset Alzheimer's disease family study. J Alzheimers Dis 2011;23:249–55
123. Mez J, Mukherjee S, Thornton T, et al. The executive prominent/memory prominent spectrum in Alzheimer's disease is highly heritable. Neurobiol Aging 2016;41:115–21
124. Beecham GW, Hamilton K, Naj AC, et al. Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. PLoS Genet 2014;10:e1004606
125. Farfel JM, Yu L, Buchman AS, et al. Relation of genomic variants for Alzheimer disease dementia to common neuropathologies. Neurology 2016;87:489–96
126. Tropf FC, Lee SH, Verweij RM, et al. Hidden heritability due to heterogeneity across seven populations. Nat Hum Behav 2017;1:757–65
127. Liu CC, Liu CC, Kanekiyo T, et al. Apolipoprotein E and Alzheimer disease: Risk, mechanisms and therapy. Nat Rev Neurol 2013;9:106–18
128. Farrer LA, Cupples LA, Haines JL, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. Jama 1997;278:1349–56
129. Ward A, Crean S, Mercaldi CJ, et al. Prevalence of apolipoprotein E4 genotype and homozygotes (APOE e4/4) among patients diagnosed with Alzheimer's disease: A systematic review and meta-analysis. Neuroepidemiology 2012;38:1–17
130. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261:921–3
131. Howie B, Fuchsberger C, Stephens M, et al. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 2012;44:955–9
132. van Leeuwen EM, Kanterakis A, Deelen P, et al. Population-specific genotype imputations using minimac or IMPUTE2. Nat Protoc 2015;10:1285–96
133. Chaudhury S, Patel T, Barber IS, et al. Polygenic risk score in postmortem diagnosed sporadic early-onset Alzheimer's disease. Neurobiol Aging 2018;62:244.e1–8
134. Escott-Price V, Bellenguez C, Wang LS, et al. Gene-wide analysis detects two new susceptibility genes for Alzheimer's disease. PLoS One 2014;9:e94661
135. Naj AC, Jun G, Reitz C, et al. Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: A genome-wide association study. JAMA Neurol 2014;71:1394–404
136. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. Jama 2010;303:1832–40
137. Bennett DA, Schneider JA, Aggarwal NT, et al. Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. Neuroepidemiology 2006;27:169–76
138. Mok W, Chow TW, Zheng L, et al. Clinicopathological concordance of dementia diagnoses by community versus tertiary care clinicians. Am J Alzheimers Dis Other Demen 2004;19:161–5
139. Schneider JA, Aggarwal NT, Barnes L, et al. The neuropathology of older persons with and without dementia from community versus clinic cohorts. J Alzheimers Dis 2009;18:691–701
140. Brayne C. Clinicopathological studies of the dementias from an epidemiological viewpoint. Br J Psychiatry 1993;162:439–46
141. Slooter AJ, Cruts M, Kalmijn S, et al. Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: The Rotterdam Study. Arch Neurol 1998;55:964–8
142. Leonenko G, Sims R, Shoai M, et al. Polygenic risk and hazard scores for Alzheimer's disease prediction. Ann Clin Transl Neurol 2019;6:456–65
143. Escott-Price V, Baker E, Shoai M, et al. Genetic analysis suggests high misassignment rates in clinical Alzheimer's cases and controls. Neurobiol Aging 2019;77:178–82
144. Escott-Price V, Myers A, Huentelman M, et al. Polygenic risk score analysis of Alzheimer's disease in cases without APOE4 or APOE2 alleles. J Prev Alzheimers Dis 2019;6:16–9
145. Escott-Price V, Myers AJ, Huentelman M, et al. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. Ann Neurol 2017;82:311–4
146. Escott-Price V, Shoai M, Pither R, et al. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. Neurobiol Aging 2017;49:214.e7–11
147. Escott-Price V, Sims R, Bannister C, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. Brain 2015;138:3673–84
148. Porter T, Burnham SC, Savage G, et al. A polygenic risk score derived from episodic memory weighted genetic variants is associated with cognitive decline in preclinical Alzheimer's disease. Front Aging Neurosci 2018;10:423
149. Tan CH, Fan CC, Mormino EC, et al. Polygenic hazard score: An enrichment marker for Alzheimer's associated amyloid and tau deposition. Acta Neuropathol 2018;135:85–93
150. Desikan RS, Fan CC, Wang Y, et al. Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score. PLoS Med 2017;14:e1002258
151. Sleegers K, Roks G, Theuns J, et al. Familial clustering and genetic risk for dementia in a genetically isolated Dutch population. Brain 2004;127:1641–9
152. Chasioti D, Yan J, Nho K, et al. Progress in polygenic composite scores in Alzheimer's and other complex diseases. Trends Genet 2019;35:371–82
153. Szefer E, Lu D, Nathoo F, et al. Multivariate association between single-nucleotide polymorphisms in Alzgene linkage regions and structural changes in the brain: Discovery, refinement and validation. Stat Appl Genet Mol Biol 2017;16:349–65
154. Ebbert MT, Ridge PG, Kauwe JS. Bridging the gap between statistical and biological epistasis in Alzheimer's disease. Biomed Res Int 2015;2015:1–7
155. So HC, Gui AH, Cherny SS, et al. Evaluating the heritability explained by known susceptibility variants: A survey of ten complex diseases. Genet Epidemiol 2011;35:310–7
156. Calero M, Gomez-Ramos A, Calero O, et al. Additional mechanisms conferring genetic susceptibility to Alzheimer's disease. Front Cell Neurosci 2015;9:138
157. Nazarian A, Kulminski AM. Evaluation of the genetic variance of Alzheimer's disease explained by the disease-associated chromosomal regions. J Alzheimers Dis 2019;70:907–15
158. Raghavan N, Tosto G. Genetics of Alzheimer's disease: The importance of polygenic and epistatic components. Curr Neurol Neurosci Rep 2017;17:78
159. Walter K, Min JL, Huang J, et al. The UK10K project identifies rare variants in health and disease. Nature 2015;526:82–90
160. Zheng HF, Ladouceur M, Greenwood CM, et al. Effect of genome-wide genotyping and reference panels on rare variants imputation. J Genet Genomics 2012;39:545–50
161. Pasaniuc B, Rohland N, McLaren PJ, et al. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. Nat Genet 2012;44:631–5

162. De Jager PL. The era of GWAS is over – No. Mult Scler 2018;24: 258–60
163. Naj AC, Schellenberg GD, Alzheimer's Disease Genetics Consortium (ADGC). Alzheimer's disease genetics C. Genomic variants, genes, and pathways of Alzheimer's disease: An overview. Am J Med Genet B Genet 2017;174:5–26
164. Visscher PM, Medland SE, Ferreira MA, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet 2006;2:e41
165. Silventoinen K, Sammalisto S, Perola M, et al. Heritability of adult body height: A comparative study of twin cohorts in eight countries. Twin Res 2003;6:399–408
166. Silventoinen K. Determinants of variation in adult body height. J Biosoc Sci 2003;35:263–85
167. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet 2014;46:1173–86
168. Lango Allen H, Estrada K, Lettre G, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 2010;467:832–8
169. Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. Nat Rev Genet 2018;19:286–98
170. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. Nat Rev Genet 2012;13:36–46
171. Ebbert MTW, Jensen TD, Jansen-West K, et al. Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. Genome Biol 2019;20:97
172. Rohilla KJ, Gagnon KT. RNA biology of disease-associated microsatellite repeat expansions. Acta Neuropathol Commun 2017;5:63
173. Loureiro JR, Oliveira CL, Silveira I. Unstable repeat expansions in neurodegenerative diseases: Nucleocytoplasmic transport emerges on the scene. Neurobiol Aging 2016;39:174–83
174. Kim JC, Mirkin SM. The balancing act of DNA repeat expansions. Curr Opin Genet Dev 2013;23:280–8
175. Ionita-Laza I, Makarov V, Consortium AAS. Scan-statistic approach identifies clusters of rare disease variants in LRP2, a gene linked and associated with autism spectrum disorders, in three datasets. Am J Hum Genet 2012;90:1002–13
176. Blue EE, Bis JC, Dorschner MO, et al. Genetic variation in genes underlying diverse dementias may explain a small proportion of cases in the Alzheimer's disease sequencing project. Dement Geriatr Cogn Disord 2018;45:1–17
177. Patel D, Mez J, Vardarajan BN, et al. Association of rare coding mutations with Alzheimer disease and other dementias among adults of European ancestry. JAMA Netw Open 2019;2:e191350
178. Bis JC, Jian X, Kunkle BW, et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. Mol Psychiatry 2018; doi: 10.1038/s41380-018-0112-7
179. Preuss C, Capredon M, Wunnemann F, et al. Family based whole exome sequencing reveals the multifaceted role of notch signaling in congenital heart disease. PLoS Genet 2016;12:e1006335
180. Maffucci P, Bigio B, Rapaport F, et al. Blacklisting variants common in private cohorts but not in public databases optimizes human exome analysis. Proc Natl Acad Sci U S A 2019;116:950–9
181. Lu RC, Yang W, Tan L, et al. Association of HLA-DRB1 polymorphism with Alzheimer's disease: A replication and meta-analysis. Oncotarget 2017;8:93219–26
182. Furukawa H, Oka S, Tsuchiya N, et al. The role of common protective alleles HLA-DRB1*13 among systemic autoimmune diseases. Genes Immun 2017;18:1–7
183. Wang ZX, Wang HF, Tan L, et al. Effects of HLA-DRB1/DQB1 genetic variants on neuroimaging in healthy, mild cognitive impairment, and Alzheimer's disease cohorts. Mol Neurobiol 2017;54:3181–8
184. Cree BA. Multiple sclerosis genetics. Handb Clin Neurol 2014;122: 193–209
185. Bertolin C, Querin G, Bozzoni V, et al. New FIG4 gene mutations causing aggressive ALS. Eur J Neurol 2018;25:e41–2
186. Osmanovic A, Rangnau I, Kosfeld A, et al. FIG4 variants in central European patients with amyotrophic lateral sclerosis: A whole-exome and targeted sequencing study. Eur J Hum Genet 2017;25:324–31
187. Chow CY, Landers JE, Bergren SK, et al. Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. Am J Hum Genet 2009;84:85–8
188. Ren Y, Reddy JS, Pottier C, et al. Identification of missing variants by combining multiple analytic pipelines. BMC Bioinformatics 2018;19: 139
189. NIAGADS. Review and proposed actions for false-positive association results in ADSP case-control data. Available at: https://www.niagads.org/adsp/content/review-and-proposed-actions-false-positive-association-results-adsp-case-control-data. Accessed December 3, 2019
190. Schmitt FA, Nelson PT, Abner E, et al. University of Kentucky Sanders-Brown healthy brain aging volunteers: Donor characteristics, procedures, and neuropathology. Curr Alzheimer Res 2012;9:724–33
191. Braak H, Braak E, Bohl J. Staging of Alzheimer-related cortical destruction. Eur Neurol 1993;33:403–8
192. Mirra SS. The CERAD neuropathology protocol and consensus recommendations for the postmortem diagnosis of Alzheimer's disease: A commentary. Neurobiol Aging 1997;18:S91–4
193. Suh E, Grando K, Van Deerlin VM. Validation of a long-read PCR assay for sensitive detection and sizing of C9orf72 hexanucleotide repeat expansions. J Mol Diagn 2018;20:871–82
194. Cleary EM, Pal S, Azam T, et al. Improved PCR based methods for detecting C9orf72 hexanucleotide repeat expansions. Mol Cell Probes 2016;30:218–24