Brain cell somatic gene recombination and its phylogenetic foundations

Received for publication, January 16, 2020, and in revised form, July 22, 2020, DOI 10.1074/jbc.REV120.009192

Gwendolyn Kaeser and Jerold Chun*

From the Degenerative Disease Program at the Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California, USA

Edited by Paul E. Fraser

A new form of somatic gene recombination (SGR) has been identified in the human brain that affects the Alzheimer's disease gene, amyloid precursor protein (*APP*). SGR occurs when a gene sequence is cut and recombined within a single cell's genomic DNA, generally independent of DNA replication and the cell cycle. The newly identified brain SGR produces genomic complementary DNAs (gencDNAs) lacking introns, which integrate into locations distinct from germline loci. This brief review will present an overview of likely related recombination mechanisms and genomic cDNA-like sequences that implicate evolutionary origins for brain SGR. Similarities and differences exist between brain SGR and VDJ recombination in the immune system, the first identified SGR form that now has a well-defined enzymatic machinery. Both require gene transcription, but brain SGR uses an RNA intermediate and reverse transcriptase (RT) activity, which are characteristics shared with endogenous retrotransposons. The identified gencDNAs have similarities to other cDNA-like sequences existing throughout phylogeny, including intron-less genes and inactive germline processed pseudogenes, with likely overlapping biosynthetic processes. gencDNAs arise somatically in an individual to produce multiple copies; can be functional; appear most frequently within postmitotic cells; have diverse sequences; change with age; and can change with disease state. Normally occurring brain SGR may represent a mechanism for gene optimization and long-term cellular memory, whereas its dysregulation could underlie multiple brain disorders and, potentially, other diseases like cancer. The involvement of RT activity implicates already Food and Drug Administration–approved RT inhibitors as possible near-term interventions for managing SGR-associated diseases and suggest next-generation therapeutics targeting SGR elements.

The human brain is the most complex organ in our body, encompassing astronomical cellular diversity, intricate connectivities, and specialized functions that include cognition and memory. It also shows vast molecular complexity underscored by the expression of half of all the protein-coding genes in the human genome and high levels of expressed gene diversity, compared with other tissues (1, 2). In the 1960s, the intricate cellular appearance of the developing goldfish brain was proposed to be analogous to Ig molecular diversity that encompasses tens of thousands of different Ig antibodies; indeed, it was hypothesized that somatic DNA sequence changes (called "copy splice" of DNA at the time) might underlie the molecular and cellular heterogeneity of the brain (3, 4). The proposed gene recombination would necessitate bringing together two or more disparate genomic sequences present within germline DNA to form a new gene sequence, thereby creating molecular heterogeneity within nongermline (*i.e.* somatic) cells. Proof of somatic gene recombination (SGR) at the Ig locus in cells of the immune system, now known as VDJ recombination, was first reported in 1976 (5) and supported the feasibility of SGR existing elsewhere. However, ensuing decades of research did not reveal SGR anywhere else, including with in the brain.

Nevertheless, hints that SGR-related processes might occur in the nervous system emerged with the identification of immune SGR machinery within the mammalian brain (6, 7). This stimulated a search for somatic genomic changes in the embryonic and postnatal brain and resulted in the discovery of pervasive genomic mosaicism, whereby single cells from the same brain vary in DNA sequence. Somatic genomic mosaicism contrasted with classical views in which all cells of an individual share the same genome. Multiple forms of genomic mosaicism were identified, including aneuploidies and aneusomies (8–11), large copy number variations (12–15), single-nucleotide variations (SNVs) (16, 17), long-interspersed nuclear element 1 (LINE1) retrotransposition (18–21), and overall DNA content variation (22–24). None of these forms, with the possible exception of rare *de novo* LINE1 retrotransposition events (18–21), produced conserved chromosomal, subchromosomal, or gene-specific genomic changes among individuals.

Possible functional roles for somatic genomic mosaicism in the brain (reviewed in Refs. 25 and 26) have been identified in animal studies through aneusomic neurons that are integrated into neuronal circuits (27), produce gene-transcription effects (28), and alter cell survival (29). These and most other assessments have linked large structural changes to functional consequences, albeit in the absence of precisely identified and affected cellular genes. The study of the functional consequences of somatic genomic mosaicism in the human brain has been especially difficult, hampered by both technical and biological challenges that include limitations of working with postmitotic cells and the human brain.

Indeed, even identification of somatic genomic mosaicism in human brain is challenging. Issues range from access to consistent, high-quality human autopsy materials of comparable age, sex, and race or sequencing technologies and bioinformatics not optimized for mosaicism to biological impediments like nonclonal genomic alterations affecting just a single postmitotic cell, obfuscated by a sea of normal cells. For example, the

* For correspondence: Jerold Chun, jchun@sbpdiscovery.org.

brain shows functional uniqueness of very small populations of cells, such as neurons responsive to input from a particular point in visual space (30, 31) or even the proposed single "grandmother" cell that responds to a complex assembly of visual features such as those comprising a specific face (32), yet most sequencing techniques only allow detection of *de novo* somatic mutations present in >1–10% of cells (33). In the human brain, this represents a background of tens of billions (10^{10}) of unaffected cells preventing mutation detection even for some clonally expanded mutations, let alone isolated single-cell changes that would require sequencing sensitivity and specificity many logs beyond current technologies. Moreover, each neuron has an anatomically unique location tied to function; without additional preparative and analytical steps (that do not readily exist for human brain studies), sequencing a desired single neuron would be broadly analogous to trying to isolate and sequence a desired single colon cancer stem cell by sequencing a large pool derived from all cells of the body: it could completely miss the targeted tissue let alone the single cell. Combined, these extreme examples illustrate how traditional approaches designed to assess germline changes in bulk peripheral tissues, or even focused approaches that might detect mutations at a 0.1% level, still remain many orders of magnitude away from detecting genomic changes within single targeted neurons of the human brain and highlight the importance of good study design and technology considerations for both cell isolation and reliable sequencing.

The existence of diverse forms of somatic genomic mosaicism increased the possibility that brain SGR might exist, albeit in the absence of a specific gene candidate. However, a promising gene candidate emerged during more recent studies of genomic mosaicism in Alzheimer's disease (AD) brains: the amyloid precursor protein (*APP*) gene. This prominent AD gene has a well-known gene dosage effect, in which its increased copy number is causal for AD in rare familial AD cases and Down syndrome (trisomy 21, with three copies of *APP* via its location on chromosome 21). Intriguingly, *APP* copy numbers were found to be *mosaically* increased in sporadic AD neuronal genomes, with evidence for amplification in discrete locations within a nucleus (22, 34). Analyses were thus focused on *APP* genomic sequences in human neurons: remarkably, the identified *APP* sequences resembled complementary DNAs (cDNAs): they lacked introns yet were present in genomic DNA and were therefore termed "gencDNAs" (Fig. 1). Their many forms and copies represented novel components of the somatic human genome yet displayed features that were highly reminiscent of known, genetic (for single-celled organisms), or germline evolutionary elements lacking introns: intron-less genes and inactive processed pseudogenes.

This review will compare and contrast gencDNAs produced by brain SGR to other phylogenic genomic elements showing gene intron loss and cDNA-like sequences. It will also broadly compare and contrast the two known types of SGR: the newly discovered brain SGR that results in gencDNA products and VDJ recombination in the immune system (in-depth reviews of VDJ recombination have been published elsewhere (35, 36)). Possible mechanisms producing brain SGR along with potential roles in aging and disease are also discussed. Other forms of DNA recombination, such as DNA repair mechanisms and meiotic recombination, are beyond the scope of this review.

SGR in the brain

SGR of *APP* manifested as many genomic *APP* variants present within a single human brain (34), forming a genomic mosaic of distinct sequences and copy numbers that varied among brain cell genomes. *APP* gencDNA sequences lacked introns, showed the basic structure of a cDNA, and included known, brain-specific RNA splice variants, *APP-751* and *APP-695* (Fig. 1). The complete *APP-770* form (that contains all exons) is generally not expressed in the brain and was not identified as a neuronal gencDNA. The presence of brain-specific RNA splice variant sequences as gencDNAs supports a requirement for an expressed RNA intermediate template in gencDNA formation, as well as reverse transcription of the RNA molecule.

Although full-length forms were observed and enriched in nondiseased brains, most gencDNAs that were prevalent in AD brains showed truncated *APP* sequences containing intraexonic junctions (IEJs), in which noncontiguous exons are fused by intraexonic short microhomology sequences, ranging from 2 to 20 base pairs. When sequenced as single, intact molecules using long-read Pacific Biosciences single-molecule real-time circular consensus sequencing, gencDNA sequence diversity included not only IEJs but numerous small insertions, deletions, and, most commonly, SNVs. A marked shift in AD *versus* nondiseased gencDNA forms and representation was observed, involving an AD-related 10-fold increase in gencDNA diversity showing a preponderance of IEJ-containing truncated gencDNAS with many insertions, deletions, and SNVs, which contrasted with a majority of full-length sequences in nondiseased neurons. Interestingly, sporadic AD neurons contained gencDNAs with 11 known *familial* AD-causing SNVs that nonetheless occurred somatically. The role for gencDNAs in disease was further supported by DNA *in situ* hybridization that revealed from 0 to 13 gencDNA loci within a single neuronal nucleus and a much higher prevalence in diseased tissue. Notably, 30–40% of AD and 75% of nondiseased neuronal nuclei showed no evidence of *APP* gencDNAs, emphasizing the importance of accounting for genomic mosaicism in experimental procedures on human brain.

The existence of *APP* gencDNAs has been independently confirmed (37), and reanalysis of published data identified gencDNA insertion sites distributed throughout the genome and away from germline *APP* on chromosome 21 (38). These findings are consistent with 1) DNA *in situ* hybridization signals that showed gencDNAs in nuclear locations that were distinct from the wildtype (WT) locus (34) and 2) the previously reported increased DNA content variation observed in human cortical neurons, that is further increased in AD (22–24) through the formation of *new* DNA sequences that are distinct from WT loci and germline DNA.

Brain SGR undoubtedly affects other genes, and indeed, IEJs in other genes were found in an independent, commercially produced, and publicly available long-read whole-exome RNA-sequencing data set (38, 39). However, SGR does appear to be

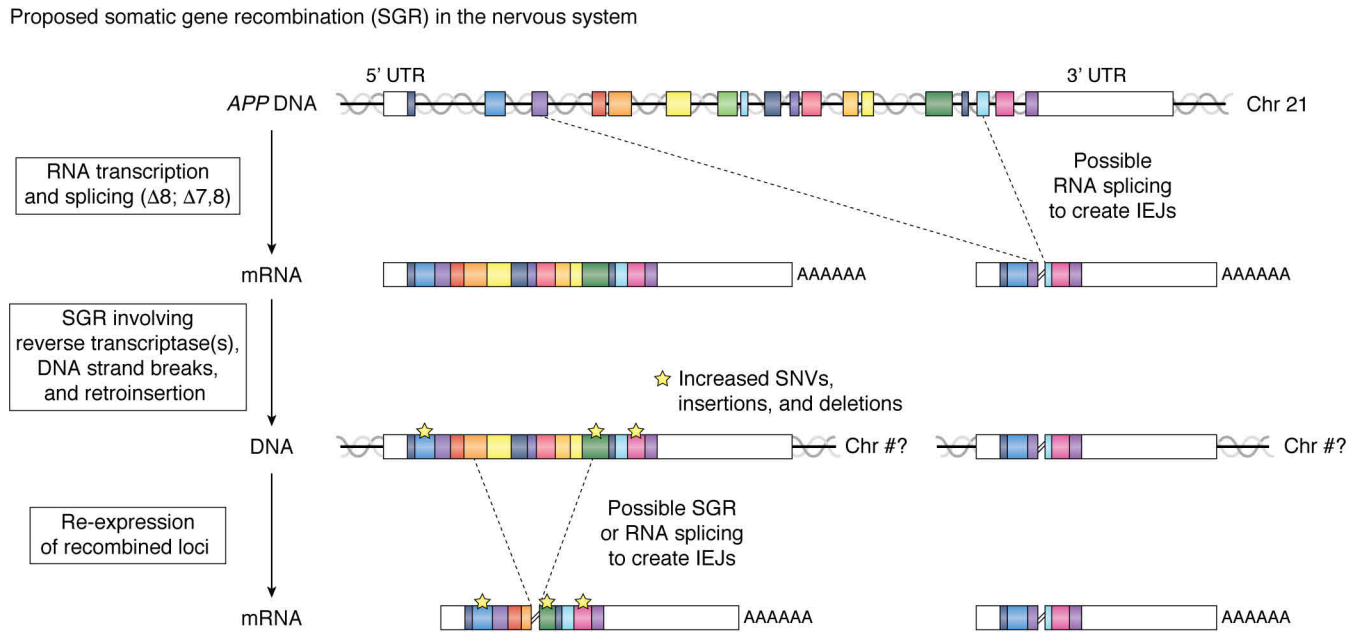

Figure 1. Structure of a gencDNA. The *APP* genomic locus on chromosome 21 (*A*) and full-length mRNA (*APP-770*; containing all 18 exons) (*B*) are provided as references. Two types of gencDNAs were identified in both RNA and DNA: full-length brain-specific isoforms (*APP-751* and *APP-695*) (*C*) and truncated sequences with IEJs (*D*). The gencDNA 3/16, which joins the middle of exon 3 with the middle of exon 16 (*i.e.* 3/16), and 1/14 are shown as examples. Known pathogenic SNVs (11 thus far) were also identified in some variants, including the in-frame gencDNAs *APP-751* and 3/16 (in-frame examples include Australian, London, French, Indiana, and A673V; shown in *red* with *yellow stars*). *Chr*, chromosome.

specific to particular genes at least in AD neurons from the prefrontal cortex: notably, *PSEN1*, another causal AD gene in rare families, was not similarly affected by SGR (34).

The proof of concept for biological repercussions of gencDNAs already has enormous support through the widespread generation and use of exogenous cDNA transgenes, in cells and animals, which are similar in structure to gencDNAs. These are most often driven by non-WT promoters, which are routinely used for bacterial or yeast transformation, mammalian cell line transfection, retrovirally mediated cDNA transduction, and production of transgenic mice, including the study of genomically mosaic effects of transgene integration and/or expression in subsets of cells. *In vitro* experimentation provides further proof of concept for gencDNA functionality: heterologous expression of in-frame coding sequences were translated into proteins and resulted in decreased cell survival. Additional *in vitro* experiments recreated gencDNAs in cell lines and identified three conditions required for SGR of *APP*: 1) *APP* transcription and splicing, 2) DNA strand breaks, and 3) reverse-transcriptase (RT) activity (Fig. 2). A defined enzymology is currently unclear and will be discussed below in relation to other conditions of intron loss, along with SGR in the immune system.

Immune SGR: VDJ recombination

VDJ recombination, which produces the antibody and T-cell receptor diversity that is integral to the adaptive immune systems of all jawed vertebrates (40), was the first confirmed type of SGR and exclusively operates in B and T cells (36, 41, 42). VDJ recombination is the process by which single variable (V), joining (J) (VJ joining in Ig light chains), and intervening diversifying (D) (VDJ joining in Ig heavy chains and T-cell receptors) gene segments recombine in genomic DNA to form the coding sequence for the antigen-binding portion of an antibody (or T-cell receptor) chain (Fig. 3). DNA cleavage is initiated by the recombination-activating genes (RAGs), *RAG1* and *RAG2*, which introduce site-specific double-strand breaks at the border between *cis* recombination signal sequences (RSSs) that flank the V, D, and J coding gene segments. The V, D, and J gene segments all have defined chromosomal locations that are maintained after recombination (*e.g.* on chromosomes 14 (heavy chain), 2, and 22 (light chains)). Cryptic RSS motifs also exist throughout the genome (43–45), some of which can result in pathogenic chromosomal translocations (35, 46, 47). Ig genes additionally undergo class-switch recombination that alters the heavy-chain constant region to produce isotype changes that employ a distinct enzymology and *cis* DNA elements (48–50). Notably, unlike brain SGR, immune SGR generally does not introduce *new* DNA content into the genome but rather reduces it. Thus, the recombined immune coding sequences are generally conserved (aside from somatic hypermutation of hypervariable regions during affinity maturation (49)) by use of the existing gene segment repertoire to create the recombined sequence that remains on the same chromosome and again produces a net loss of DNA by removing intervening genomic DNA during recombination.

cDNA-like genomic sequences are present throughout phylogeny

Intron loss—Brain SGR produces mosaic gene copies that lack introns. This is almost certainly due to the activity of RTs acting on spliced mRNAs. Although perhaps surprising in the context of the human brain, varied intron prevalence is found across phylogeny and has functional implications. Ninety-four

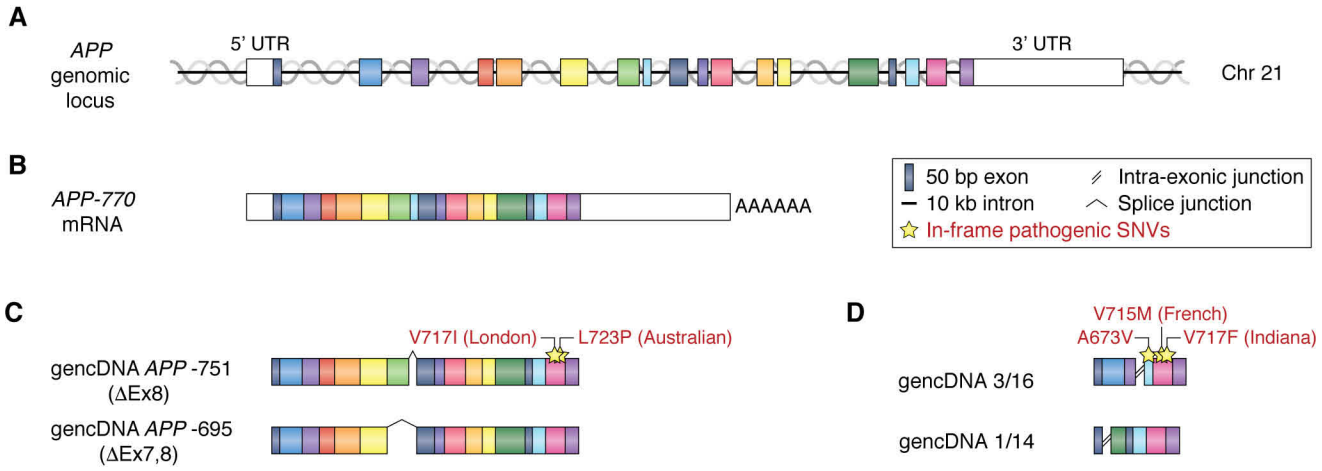

Figure 2. Proposed mechanism of SGR in the nervous system. SGR in the nervous system (*APP* locus shown as an example) initially begins at the WT genomic locus, which is transcribed into RNA and spliced into mRNA. Reverse transcriptases then retroinsert cDNA-like sequences into regions of broken DNA that occur at multiple locations in the genome (*Chr #?*). This process is concomitant with the introduction of SNVs, insertions, and deletions, particularly by reverse transcriptase that shows reduced copying fidelity. New genomic loci may be re-expressed, potentially via novel promoters associated with the insertion site, generating new mRNAs that do not require splicing. IEJs may be introduced by SGR in genomic DNA or through RNA splicing at multiple steps. *Chr*, chromosome.

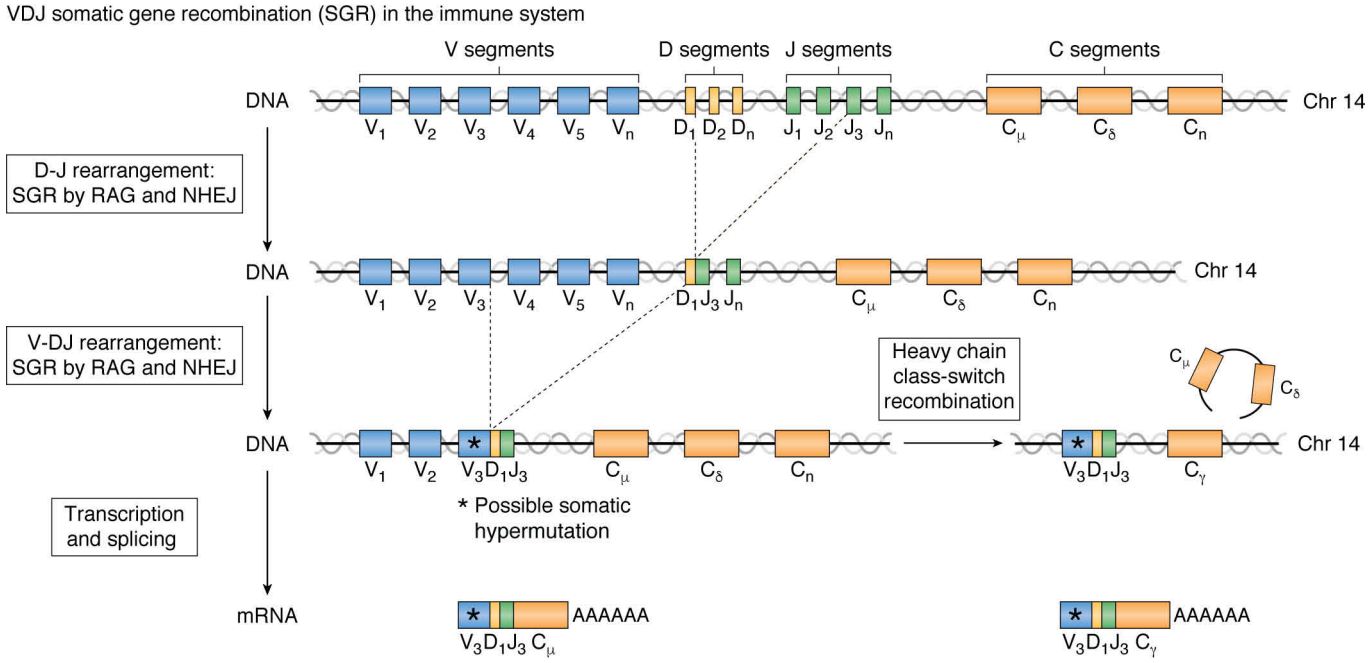

Figure 3. VDJ recombination in the immune system. VDJ recombination requires the RAG complex and NHEJ enzymes to break and recombine the native genomic locus in a multistep process. Somatic hypermutation may occur. Immunoglobulin heavy-chain class-switch recombination results in isotype replacement of the expressed C region for another downstream C region involving distinct *cis* elements and enzymology. Recombined loci are then transcribed and spliced for translation into antibody proteins. Related VDJ processes affect immunoglobulin light chains (producing VJ joining) and T-cell receptors. *Chr*, chromosome.

percent of human genes contain introns, averaging seven per gene, many of which span over 10 kb (51, 52). Some introns are functional and contain elements involved in transcriptional regulation, splicing, or genome organization or contain small nested genes such as long noncoding RNAs or microRNAs (53). Comparatively, prokaryotes and early eukaryotes are thought to have once had intron-rich genomes, but now have—through genome streamlining and compaction—few to no introns. For example, *Saccharomyces cerevisiae* contains ~290 introns in only 283 of its roughly 6,200 genes (fewer than 5%) (54–56), which average just 100–400 bp in length (56). The evolutionary history of eukaryotes has been dominated by intron loss and resulted in fewer introns within the genes of modern single-celled eukaryotes such as yeast (*S. cerevisiae*) (57). Reports of systematic intron deletion in *S. cerevisiae* to study the functional consequences of intron loss on growth and survival have revealed regulatory roles for introns in gene expression (58), although deletion is not necessarily tied to survival, and a majority of introns could be removed without terminal consequences (59). Examples of intron loss also exist in

mitochondrial ancestors (60, 61) and mammals including mice and humans (62). Intron loss could have kinetic and energetic advantages for cellular function (53, 63), including more efficient transcription and splice signals (determined through conservation of splice sites), when compared with intron-rich genomes that tend to have weak, error-prone splice signals (64). Altogether, these data suggest that brain SGR may act to improve efficiency of frequently transcribed and spliced genes through intronic loss to produce gencDNAs.

The mechanisms of intron loss are largely unknown but include reverse transcription (65–69). For example, yeast contain an active RT encoded within the Ty1 retrotransposon (70) that uses a "copy–paste" mechanism, whereby its mRNA is reverse-transcribed into cDNA and can be retrotransposed back into the genome. However, yeast contain very few processed pseudogenes (discussed in more detail below), suggesting that, over time, yeast may have replaced existing genomic loci via preferential homologous recombination (67, 71, 72). An RNA-mediated mechanism is supported by precise exon deletion, 3′ bias, and combined excision of adjacent exons (69, 73). The yeast gene *RAD52* mediates cellular gene homologous replacement (71, 74), and interestingly, a similar RNA-mediated process involving mammalian *RAD52* has been reported as a DNA repair mechanism in postmitotic neurons (75); however, a relationship to reverse transcription remains to be determined. These mechanistic hints from yeast support the concept that similar mechanisms of intron loss, such as human retrotransposons or *RAD52* and related genes, may be involved in brain SGR of *APP* and possibly other gene loci.

Processed pseudogenes—Metazoan genomes are characterized by genes that mostly contain introns. However, there are cDNA-like sequences that are proven germline elements, known as "processed pseudogenes" (76, 77). Processed pseudogenes are conventionally characterized as inactive or nonfunctional copies of existing functional genes. They lack introns and, critically, are generationally stable in sequence and genomic location. They also show reduced sequence homology at both ends, compared with the WT gene, and typically contain poly(A) tracts and short direct repeats of varying lengths. gencDNAs of *APP* show some similarities to processed pseudogenes yet also key differences (Table 1). Processed pseudogenes and gencDNAs both lack introns and show reduced sequence homology produced by SNVs. Processed pseudogenes exist within the germline at precise locations that are stable during organismal development and mitotic cell proliferation and are classically defined as being inactive. IEJs have not been reported in either processed pseudogene DNA or, in cases where they are transcribed, RNA. In clear contrast, gencDNAs are somatic (not in the germline), can occur in postmitotic cells, have myriad forms and copies within a single brain, can produce multiple and varied copies from cell to cell, integrate in diverse genomic locations, and encode open reading frames that can be translated with functional consequences for cell survival (34). Notably, although two known *APP* mRNA isoform sequences were identified as gencDNAs, at least 40 different exon combinations with previously unknown IEJs were also identified. Most differences observed between processed pseudogenes and gencDNAs support the latter as being a distinct genomic and functional entity.

The mechanisms that produce both processed pseudogenes and gencDNAs appear to include a requirement for reverse transcription. Processed pseudogenes are present throughout eukaryotic genomes and range from a few in yeast and *Drosophila* to many (thousands) in mammals. Retrotransposable elements like the Ty1 element in yeast and LINE1 in mammals are thought to participate in the generation of pseudogenes, consistent with an RT-mediated mechanism. Interestingly, WT gene intron loss is highest in parental genes that also have processed pseudogene counterparts (65): there is a positive correlation between the number of processed pseudogenes and the level of intron loss, possibly representing an interplay between WT and cDNA-like copies of a gene. Additionally, in mammals, intron loss occurs more frequently in the most highly expressed genes, such as housekeeping genes (62), which might reflect expression pressures favoring gene activity.

The activity level of processed pseudogene generation in the germline is considered sufficient enough to promote speciation, and once established, processed pseudogenes become generationally stable, thereby defining certain characteristics of a given species. By contrast, gencDNA production occurs somatically and appears to be far more active than both germline-processed pseudogene and "somatic processed pseudogenes," which have been described in proliferating cancer cells (78, 79), compared with the numbers and forms of gencDNAs thus far documented in even a single brain. Importantly, gencDNAs contrast with cancer somatic processed pseudogenes by occurring in postmitotic cells, producing multiple copies per cell and producing vast numbers of diverse gencDNAs, within a single tissue.

Possible machinery for brain SGR—The enzymology required for VDJ recombination and retrotransposition through mobile elements is generally well-characterized, whereas comparatively little is known about the enzymology of brain SGR (Table 1). *In vitro* experiments have demonstrated that both processes require DNA strand breaks and transcription, but brain SGR also depends on RT activity. In VDJ recombination, DNA double-strand breaks are produced by RAG1 and 2, after which nonhomologous end-joining (NHEJ) DNA repair molecules (Ku70, Ku80, DNA-Pkcs, XRCC4, and DNA ligase IV) join the V, D, and J gene segments (Fig. 3). However, the processes that create and repair DNA breaks in brain SGR are currently unknown but could involve endogenous LINE1 retrotransposons that encode for an endonuclease. Alternatively, they may utilize existing DNA repair mechanisms, such as NHEJ and microhomology-mediated end joining (80, 81). Notably, SGR of *APP* occurs away from its WT locus, unlike VDJ recombination, which recombines within the germline loci using conserved *cis* motifs (RSS). No conserved sequence has been identified for brain SGR, although some *cis* motifs might exist and represent an area for future study.

Both VDJ recombination and brain SGR are tightly linked to gene transcription (34, 41), suggesting a requirement for open and accessible chromatin. However, brain SGR additionally requires an RNA intermediate template and RT activity, followed by retro-insertion of the gencDNA sequence into the

Table 1
Defining characteristics of SGR of the nervous system and immune system (VDJ recombination) and retrotransposition of germline processed pseudogenes

	SGR	VDJ recombination	Retrotransposition of germline processed pseudogenes
Mechanistic requirements			
DNA strand breaks; mechanism	Yes; unknown	Yes; RAGs	Yes; retro-elements
Transcription	Yes	Yes	Yes
Reverse-transcriptase activity	Yes	No	Yes
RNA intermediate	Yes	No	Yes
DNA repair process; mechanism	Likely; unknown	Yes; NHEJ	Likely; multiple
cis-Signal sequence	Unknown	Yes; RSS	No
Cell cycle/cell division	No	Yes	Unknown
Observed characteristics			
Actively transcribed	Yes (*in vitro*)	Yes	Variable
Number of recombined loci/cell	13 observed thus far	1	NA[a]
Intraexonic junctions	Yes	No	None reported
Postmitotic somatic insertion	Yes	No	No
Functional effects	*APP* coding diversity, insertional effects, and cell survival	Ig and T-cell receptor diversity	By definition, inactive
Genomic location	Non-wildtype loci; diverse	Wildtype locus	Constant, heritable genomic locus; non-wildtype

[a]NA, not applicable.

genome, a process that is not part of VDJ recombination. Multiple findings support the need for an RNA intermediate and RT activity. First, identical sequences are present in both mRNA and genomic DNA, including brain-specific isoforms lacking defined exons and produced by RNA splicing. Second, gencDNA loci are distinctly located away from WT genomic loci (34) and integrated into multiple genomic sites (38). Third, the SNVs are consistent with the actions of an RT that has markedly reduced copying fidelity compared with DNA polymerases and is known to generate SNV mutations (*e.g.* HIV RT shows reduced copying fidelity through which HIV acquires drug resistance).

Nearly half of the human genome is attributed to the actions and remnant sequences of transposons, including the LINE1 retrotransposons (82) and human endogenous retroviruses (83). A vast majority of these remnant sequences are thought to be inactive, like pseudogenes. Similarly, LINE1 elements have been proposed as a major driver of human genome diversification by producing processed pseudogenes throughout evolution (84–86), which highlights functions distinct from self-retrotransposition. A distinct function for LINE1, human endogenous retroviruses, and other retrotransposable elements might be to provide the RT (and perhaps other) activity to support brain SGR. In this scenario, RNA expression of these elements would be less for self-retrotransposition—which could still occur, if infrequently—but rather would provide RT activity (and perhaps endonuclease and related activities) to promote gencDNA formation from cellular mRNAs. Such a process could lead to novel gene expression and protein functions, as has been reported in model systems studying the yeast Ty1 and human LINE1 retrotransposons (67, 74, 84). Mosaic *de novo* LINE1 insertions also exist and are themselves another form of brain genomic mosaicism (18–21, 87), with estimates ranging from 0.6 per genome (87) to ~14 per genome (21). Possible effects of somatic LINE1 expression and retrotransposition in neurological disorders was recently reviewed (88); relationships to SGR and gencDNAs remain to be explored.

Lastly, the discovery of RAG1 and other VDJ recombination machinery in the mammalian brain (6, 7) suggest that they may also play a role in brain SGR. Interestingly, the RAG proteins, and likely the RSSs they recognize, evolved from an ancient DNA transposase ~500 million years ago (89–92), and a fully active RAG transposon was recently identified in lancelets that are thought to be the common chordate ancestor (93). It is possible that human RAG1 and RAG2 have retained some of these activities, based upon *in vitro* assays revealing RAG transposition (94–96). Therefore, the unknown role of expressed RAG1 in the brain (6, 97), along with many other transposases (98), could be relevant to brain SGR through related activities that modulate gene transcription, DNA strand breaks, and/or reverse transcription required for brain SGR.

SGR in age and disease

The most common risk factor for neurodegenerative disease is age. Multiple forms of somatic genomic mosaicism have been shown to increase in disease, including brain SGR of *APP* (34) and somatic SNVs (17, 99). Interestingly, retrotransposon expression increases with age (reviewed in Ref. 100), including in neurons (101), where retrotransposons are already thought to exist in higher numbers (102). Increased retrotransposon RNA expression correlates with cellular age in diverse eukaryotic organisms (101, 103–105) and is proposed to produce retrotransposition; however, the formal demonstration of active, endogenous element integration *in vivo* remains under in-vestigation. The potential age-related increases in SGR and gencDNAs could explain the decades required for neurodegenerative disease to manifest, as well as the common comorbidity of two or more neurodegenerative diseases (106), whereby a threshold of SGR activity operates on distinct sets of pathogenic genes; as pathogenic gene variants accumulate over time, disease initiation results. Such an age-dependent mechanism would also predict the operation of SGR in familial forms of brain disease where, despite carrying one or more disease-promoting germline mutations, decades (albeit fewer) of life are still required for disease manifestation, as seen in familial and Down syndrome AD (107). Similarly, SGR actions on multiple yet distinct sets of known and unknown pathogenic genes could promote comorbidity with other diseases as observed

commonly with AD (*e.g.* Parkinson's disease) (106, 108). Dysregulated SGR in brain disorders like AD implicates a range of SGR-related processes involving molecular targets that could be therapeutically accessible by targeting variant genes, their expressed products or specific enzymes—for example, by using the RT inhibitors that are already approved by the Food and Drug Administration for treating HIV (106, 109); *post hoc* literature analyses support potential benefits of RTs in AD (106, 109–111).

Conclusions

Brain SGR represents a new mechanism for diversifying gene expression, through the "recording" and "playback" of new gene forms within single somatic-cell genomes of an individual, particularly in long-lived postmitotic neurons. It can optimize and "hard-wire" complex and plastic germline genomes containing intronic genes by conversion to streamlined gene copies that lack introns to complement WT alleles. Hard-wiring selected genomes with intron-less gencDNAs is analogous to evolutionarily produced gene organization in unicellular prokaryotes and eukaryotes, which generally lack introns to access genomic and energetic efficiencies, albeit at the expense of plasticity. Previously, the concept of "ribotype" was proposed (112, 113), whereby the spectrum of RNAs emerging from the WT intron-containing genome could be supplemented or replaced over time via a mechanism that memorializes a preferred ribotype within the genome; gencDNAs can provide such a mechanism. This form of genomic streamlining likely involves shared elements of retrotransposition that have genomically shaped humans and all other species through evolutionary intronic changes (69). SGR in the brain appears to access these or similar processes to endow it with a somatic mechanism that could underlie many brain phenomena and functions—not only well-known cellular diversity but also neural plasticity and long-term "memory" storage—through the production of gencDNAs. Indeed, a principle feature of the brain is long-term memory, and the concept that memory might in some way be encoded within DNA has been suggested (114, 115). SGR might thus create the stable storage of information linked to neural activity, to enable genomic sculpting, over time, of the brain's functions.

Acknowledgments—We thank Danielle Jones and Dr. Laura Wolszon for editorial assistance and Dr. David Schatz for helpful discussions.

Author contributions—G. K. and J. C. conceptualization; G. K. and J. C. writing-original draft; G. K. and J. C. writing-review and editing; J. C. resources; J. C. supervision; J. C. funding acquisition.

Funding and additional information—This work was supported by NIA, National Institutes of Health Grants R56AG067489 and R01AG065541 (to J. C.). This work was also supported by nonfederal funds from the Shaffer Family Foundation, the Bruce Ford & Anne Smith Bundy Foundation, and the Hervey Family Fund (to J. C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—J. C. is a co-founder of Mosaic Pharmaceuticals.

Abbreviations—The abbreviations used are: SGR, somatic gene recombination; RAG, recombination-activating gene; IEJ, intraexonic junction; NHEJ, nonhomologous end joining; RSS, recombination signal sequence; APP, amyloid precursor protein; AD, Alzheimer's disease; gencDNA, genomic complementary DNA; RT, reverse transcriptase; SNV, single-nucleotide variation; LINE1, long-interspersed nuclear element 1; cDNA, complementary DNA.

References

1. Naumova, O. Y., Lee, M., Rychkov, S. Y., Vlasova, N. V., and Grigorenko, E. L. (2013) Gene expression in the human brain: the current state of the study of specificity and spatiotemporal dynamics. *Child Dev.* **84,** 76–88 CrossRef Medline
2. Colantuoni, C., Purcell, A. E., Bouton, C. M., and Pevsner, J. (2000) High throughput analysis of gene expression in the human brain. *J. Neurosci. Res.* **59,** 1–10 CrossRef Medline
3. Dreyer, W., Gray, W., and Hood, L. (1967) The genetic, molecular, and cellular basis of antibody formation: some facts and a unifying hypothesis. *Cold Spring Harb. Symp. Quant. Biol.* **32,** 353–367 CrossRef
4. Dreyer, W. J., and Bennett, J. C. (1965) The molecular basis of antibody formation: a paradox. *Proc. Natl. Acad. Sci. U.S.A.* **54,** 864–869 CrossRef Medline
5. Hozumi, N., and Tonegawa, S. (1976) Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. U.S.A.* **73,** 3628–3632 CrossRef Medline
6. Chun, J. J., Schatz, D. G., Oettinger, M. A., Jaenisch, R., and Baltimore, D. (1991) The recombination activating gene-1 (RAG-1) transcript is present in the murine central nervous system. *Cell* **64,** 189–200 CrossRef Medline
7. Gao, Y., Sun, Y., Frank, K. M., Dikkes, P., Fujiwara, Y., Seidl, K. J., Sekiguchi, J. M., Rathbun, G. A., Swat, W., Wang, J., Bronson, R. T., Malynn, B. A., Bryans, M., Zhu, C., Chaudhuri, J., *et al.* (1998) A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95,** 891–902 CrossRef Medline
8. Rehen, S. K., Yung, Y. C., McCreight, M. P., Kaushal, D., Yang, A. H., Almeida, B. S., Kingsbury, M. A., Cabral, K. M., McConnell, M. J., Anliker, B., Fontanoz, M., and Chun, J. (2005) Constitutional aneuploidy in the normal human brain. *J. Neurosci.* **25,** 2176–2180 CrossRef Medline
9. Rehen, S. K., McConnell, M. J., Kaushal, D., Kingsbury, M. A., Yang, A. H., and Chun, J. (2001) Chromosomal variation in neurons of the developing and adult mammalian nervous system. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 13361–13366 CrossRef Medline
10. Westra, J. W., Peterson, S. E., Yung, Y. C., Mutoh, T., Barral, S., and Chun, J. (2008) Aneuploid mosaicism in the developing and adult cerebellar cortex. *J. Comp. Neurol.* **507,** 1944–1951 CrossRef Medline
11. Yurov, Y. B., Iourov, I. Y., Monakhov, V. V., Soloviev, I. V., Vostrikov, V. M., and Vorsanova, S. G. (2005) The variation of aneuploidy frequency in the developing and adult human brain revealed by an interphase FISH study. *J. Histochem. Cytochem.* **53,** 385–390 CrossRef Medline
12. Cai, X., Evrony, G. D., Lehmann, H. S., Elhosary, P. C., Mehta, B. K., Poduri, A., and Walsh, C. A. (2014) Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep.* **8,** 1280–1289 CrossRef Medline
13. McConnell, M. J., Lindberg, M. R., Brennand, K. J., Piper, J. C., Voet, T., Cowing-Zitron, C., Shumilina, S., Lasken, R. S., Vermeesch, J. R., Hall, I. M., and Gage, F. H. (2013) Mosaic copy number variation in human neurons. *Science* **342,** 632–637 CrossRef Medline
14. Gole, J., Gore, A., Richards, A., Chiu, Y. J., Fung, H. L., Bushman, D., Chiang, H. I., Chun, J., Lo, Y. H., and Zhang, K. (2013) Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. *Nat. Biotechnol.* **31,** 1126–1132 CrossRef Medline
15. Rohrback, S., April, C., Kaper, F., Rivera, R. R., Liu, C. S., Siddoway, B., and Chun, J. (2018) Submegabase copy number variations arise during

cerebral cortical neurogenesis as revealed by single-cell whole-genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **115,** 10804–10809 CrossRef Medline
16. Evrony, G. D., Lee, E., Park, P. J., and Walsh, C. A. (2016) Resolving rates of mutation in the brain using single-neuron genomics. *eLife* **5,** e12966 CrossRef Medline
17. Lodato, M. A., Rodin, R. E., Bohrson, C. L., Coulter, M. E., Barton, A. R., Kwon, M., Sherman, M. A., Vitzthum, C. M., Luquette, L. J., Yandava, C. N., Yang, P., Chittenden, T. W., Hatem, N. E., Ryu, S. C., Woodworth, M. B., *et al.* (2018) Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359,** 555–559 CrossRef Medline
18. Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., Morell, M., O'Shea, K. S., Moran, J. V., and Gage, F. H. (2009) L1 retrotransposition in human neural progenitor cells. *Nature* **460,** 1127–1131 CrossRef Medline
19. Muotri, A. R., Chu, V. T., Marchetto, M. C., Deng, W., Moran, J. V., and Gage, F. H. (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435,** 903–910 CrossRef Medline
20. Baillie, J. K., Barnett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapio, F., Brennan, P. M., Rizzu, P., Smith, S., Fell, M., Talbot, R. T., Gustincich, S., Freeman, T. C., Mattick, J. S., Hume, D. A., *et al.* (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479,** 534–537 CrossRef Medline
21. Upton, K. R., Gerhardt, D. J., Jesuadian, J. S., Richardson, S. R., Sánchez-Luque, F. J., Bodea, G. O., Ewing, A. D., Salvador-Palomeque, C., van der Knaap, M. S., Brennan, P. M., Vanderver, A., and Faulkner, G. J. (2015) Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* **161,** 228–239 CrossRef Medline
22. Bushman, D. M., Kaeser, G. E., Siddoway, B., Westra, J. W., Rivera, R. R., Rehen, S. K., Yung, Y. C., and Chun, J. (2015) Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *eLife* **4,** e05116 CrossRef Medline
23. Westra, J. W., Rivera, R. R., Bushman, D. M., Yung, Y. C., Peterson, S. E., Barral, S., and Chun, J. (2010) Neuronal DNA content variation (DCV) with regional and individual differences in the human brain. *J. Comp. Neurol.* **518,** 3981–4000 CrossRef Medline
24. Fischer, H. G., Morawski, M., Brückner, M. K., Mittag, A., Tarnok, A., and Arendt, T. (2012) Changes in neuronal DNA content variation in the human brain during aging. *Aging Cell* **11,** 628–633 CrossRef Medline
25. Bushman, D. M., and Chun, J. (2013) The genomically mosaic brain: aneuploidy and more in neural diversity and disease. *Semin. Cell. Dev. Biol.* **24,** 357–369 CrossRef Medline
26. Rohrback, S., Siddoway, B., Liu, C. S., and Chun, J. (2018) Genomic mosaicism in the developing and adult brain. *Dev. Neurobiol.* **78,** 1026–1048 CrossRef Medline
27. Kingsbury, M. A., Friedman, B., McConnell, M. J., Rehen, S. K., Yang, A. H., Kaushal, D., and Chun, J. (2005) Aneuploid neurons are functionally active and integrated into brain circuitry. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 6143–6147 CrossRef Medline
28. Kaushal, D., Contos, J. J., Treuner, K., Yang, A. H., Kingsbury, M. A., Rehen, S. K., McConnell, M. J., Okabe, M., Barlow, C., and Chun, J. (2003) Alteration of gene expression by chromosome loss in the postnatal mouse brain. *J. Neurosci.* **23,** 5599–5606 CrossRef Medline
29. Peterson, S. E., Yang, A. H., Bushman, D. M., Westra, J. W., Yung, Y. C., Barral, S., Mutoh, T., Rehen, S. K., and Chun, J. (2012) Aneuploid cells are differentially susceptible to caspase-mediated death during embryonic cerebral cortical development. *J. Neurosci.* **32,** 16213–16222 CrossRef Medline
30. Hecker, A., Schulze, W., Oster, J., Richter, D. O., and Schuster, S. (2020) Removing a single neuron in a vertebrate brain forever abolishes an essential behavior. *Proc. Natl. Acad. Sci. U.S.A.* **117,** 3254–3260 CrossRef Medline
31. Okuyama, T. (2018) Social memory engram in the hippocampus. *Neurosci. Res.* **129,** 17–23 CrossRef Medline
32. Gross, C. G. (2002) Genealogy of the "grandmother cell." *Neuroscientist* **8,** 512–518 CrossRef Medline
33. Nicolas, G., and Veltman, J. A. (2019) The role of *de novo* mutations in adult-onset neurodegenerative disorders. *Acta Neuropathol.* **137,** 183–207 CrossRef Medline
34. Lee, M. H., Siddoway, B., Kaeser, G. E., Segota, I., Rivera, R., Romanow, W. J., Liu, C. S., Park, C., Kennedy, G., Long, T., and Chun, J. (2018) Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **563,** 639–645 CrossRef Medline
35. Roth, D. B. (2014) V(D)J recombination: mechanism, errors, and fidelity. *Microbiol. Spectr.* **2,** CrossRef Medline
36. Schatz, D. G., and Swanson, P. C. (2011) V(D)J recombination: mechanisms of initiation. *Annu. Rev. Genet.* **45,** 167–202 CrossRef Medline
37. Park, J. S., Lee, J., Jung, E. S., Kim, M. H., Kim, I. B., Son, H., Kim, S., Kim, S., Park, Y. M., Mook-Jung, I., Yu, S. J., and Lee, J. H. (2019) Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of Tau phosphorylation. *Nat. Commun.* **10,** 3090 CrossRef Medline
38. Lee, M.-H., Liu, C. S., Zhu, Y., Kaeser, G. E., Rivera, R., Romanow, W. J., Kihara, Y., and Chun, J. (2020) Reply: Evidence that APP gene copy number changes reflect recombinant vector contamination. *bioRxiv* CrossRef
39. (2016) Data Release: Alzheimer Brain Isoform Sequencing (Iso-Seq) Dataset
40. Carmona, L. M., and Schatz, D. G. (2017) New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *FEBS J.* **284,** 1590–1605 CrossRef Medline
41. Schatz, D. G., and Ji, Y. (2011) Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* **11,** 251–263 CrossRef Medline
42. Schatz, D. G., and Baltimore, D. (2004) Uncovering the V(D)J recombinase. *Cell* **116,** S103–S106 CrossRef Medline
43. Merelli, I., Guffanti, A., Fabbri, M., Cocito, A., Furia, L., Grazini, U., Bonnal, R. J., Milanesi, L., and McBlane, F. (2010) RSSsite: a reference database and prediction tool for the identification of cryptic recombination signal sequences in human and murine genomes. *Nucleic Acids Res.* **38,** W262–W267 CrossRef Medline
44. Lee, A. I., Fugmann, S. D., Cowell, L. G., Ptaszek, L. M., Kelsoe, G., and Schatz, D. G. (2003) A functional analysis of the spacer of V(D)J recombination signal sequences. *PLoS Biol.* **1,** E1 CrossRef Medline
45. Rahman, N. S., Godderz, L. J., Stray, S. J., Capra, J. D., and Rodgers, K. K. (2006) DNA cleavage of a cryptic recombination signal sequence by RAG1 and RAG2: implications for partial V(H) gene replacement. *J. Biol. Chem.* **281,** 12370–12380 CrossRef Medline
46. Numata, M., Saito, S., and Nagata, K. (2010) RAG-dependent recombination at cryptic RSSs within TEL-AML1 t(12;21)(p13;q22) chromosomal translocation region. *Biochem. Biophys. Res. Commun.* **402,** 718–724 CrossRef Medline
47. Lieber, M. R., Yu, K., and Raghavan, S. C. (2006) Roles of nonhomologous DNA end joining, V(D)J recombination, and class switch recombination in chromosomal translocations. *DNA Repair (Amst.)* **5,** 1234–1245 CrossRef Medline
48. Dudley, D. D., Chaudhuri, J., Bassing, C. H., and Alt, F. W. (2005) Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Adv. Immunol.* **86,** 43–112 CrossRef Medline
49. Methot, S. P., and Di Noia, J. M. (2017) Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv. Immunol.* **133,** 37–87 CrossRef Medline
50. Xu, Z., Zan, H., Pone, E. J., Mai, T., and Casali, P. (2012) Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nat. Rev. Immunol.* **12,** 517–531 CrossRef Medline
51. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 CrossRef Medline
52. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., *et al.* (2001)

The sequence of the human genome. *Science* **291,** 1304–1351 CrossRef Medline
53. Chorev, M., and Carmel, L. (2012) The function of introns. *Front. Genet.* **3,** 55 CrossRef Medline
54. Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., *et al.* (1996) Life with 6000 genes. *Science* **274,** 546, 563–547 CrossRef Medline
55. Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423,** 241–254 CrossRef Medline
56. Spingola, M., Grate, L., Haussler, D., and Ares, M., Jr. (1999) Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* **5,** 221–234 CrossRef Medline
57. Csuros, M., Rogozin, I. B., and Koonin, E. V. (2011) A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS Comput. Biol.* **7,** e1002150 CrossRef Medline
58. Juneau, K., Miranda, M., Hillenmeyer, M. E., Nislow, C., and Davis, R. W. (2006) Introns regulate RNA and protein abundance in yeast. *Genetics* **174,** 511–518 CrossRef Medline
59. Parenteau, J., Durand, M., Véronneau, S., Lacombe, A. A., Morin, G., Guérin, V., Cecez, B., Gervais-Bird, J., Koh, C. S., Brunelle, D., Wellinger, R. J., Chabot, B., and Abou Elela, S. (2008) Deletion of many yeast introns reveals a minority of genes that require splicing for function. *Mol. Biol. Cell* **19,** 1932–1941 CrossRef Medline
60. Cuenca, A., Ross, T. G., Graham, S. W., Barrett, C. F., Davis, J. I., Seberg, O., and Petersen, G. (2016) Localized retroprocessing as a model of intron loss in the plant mitochondrial genome. *Genome Biol. Evol.* **8,** 2176–2189 CrossRef Medline
61. Pogoda, C. S., Keepers, K. G., Nadiadi, A. Y., Bailey, D. W., Lendemer, J. C., Tripp, E. A., and Kane, N. C. (2019) Genome streamlining via complete loss of introns has occurred multiple times in lichenized fungal mitochondria. *Ecol. Evol.* **9,** 4245–4263 CrossRef Medline
62. Coulombe-Huntington, J., and Majewski, J. (2007) Characterization of intron loss events in mammals. *Genome Res.* **17,** 23–32 CrossRef Medline
63. Lane, N., and Martin, W. (2010) The energetics of genome complexity. *Nature* **467,** 929–934 CrossRef Medline
64. Irimia, M., Penny, D., and Roy, S. W. (2007) Coevolution of genomic intron number and splice sites. *Trends Genet.* **23,** 321–325 CrossRef Medline
65. Zhu, T., and Niu, D. K. (2013) Frequency of intron loss correlates with processed pseudogene abundance: a novel strategy to test the reverse transcriptase model of intron loss. *BMC Biol.* **11,** 23 CrossRef Medline
66. Fink, G. R. (1987) Pseudogenes in yeast? *Cell* **49,** 5–6 CrossRef Medline
67. Derr, L. K., Strathern, J. N., and Garfinkel, D. J. (1991) RNA-mediated recombination in *S. cerevisiae*. *Cell* **67,** 355–364 CrossRef Medline
68. Roy, S. W., and Gilbert, W. (2005) The pattern of intron loss. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 713–718 CrossRef Medline
69. Roy, S. W., and Gilbert, W. (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.* **7,** 211–221 CrossRef Medline
70. Curcio, M. J., Lutz, S., and Lesage, P. (2014) The Ty1 LTR-retrotransposon of budding yeast, *Saccharomyces cerevisiae*. In *Mobile DNA*, 3rd Ed., pp. 927–964, American Society for Microbiology, Washington, D. C.
71. Derr, L. K. (1998) The involvement of cellular recombination and repair genes in RNA-mediated recombination in *Saccharomyces cerevisiae*. *Genetics* **148,** 937–945 Medline
72. Derr, L. K., and Strathern, J. N. (1993) A role for reverse transcripts in gene conversion. *Nature* **361,** 170–173 CrossRef Medline
73. Catania, F. (2017) From intronization to intron loss: How the interplay between mRNA-associated processes can shape the architecture and the expression of eukaryotic genes. *Int. J. Biochem. Cell Biol.* **91,** 136–144 CrossRef Medline
74. Maxwell, P. H., and Curcio, M. J. (2007) Retrosequence formation restructures the yeast genome. *Genes Dev.* **21,** 3308–3318 CrossRef Medline
75. Welty, S., Teng, Y., Liang, Z., Zhao, W., Sanders, L. H., Greenamyre, J. T., Rubio, M. E., Thathiah, A., Kodali, R., Wetzel, R., Levine, A. S., and Lan, L. (2018) RAD52 is required for RNA-templated recombination repair in post-mitotic neurons. *J. Biol. Chem.* **293,** 1353–1362 CrossRef Medline
76. Kovalenko, T. F., and Patrushev, L. I. (2018) Pseudogenes as functionally significant elements of the genome. *Biochemistry (Mosc.)* **83,** 1332–1349 CrossRef Medline
77. Balakirev, E. S., and Ayala, F. J. (2003) Pseudogenes: are they "junk" or functional DNA? *Annu. Rev. Genet.* **37,** 123–151 CrossRef Medline
78. Kazazian, H. H. Jr (2014) Processed pseudogene insertions in somatic cells. *Mob. DNA* **5,** 20 CrossRef Medline
79. Cooke, S. L., Shlien, A., Marshall, J., Pipinikas, C. P., Martincorena, I., Tubio, J. M., Li, Y., Menzies, A., Mudie, L., Ramakrishna, M., Yates, L., Davies, H., Bolli, N., Bignell, G. R., Tarpey, P. S., *et al.* (2014) Processed pseudogenes acquired somatically during cancer development. *Nat. Commun.* **5,** 3644 CrossRef Medline
80. Seol, J. H., Shim, E. Y., and Lee, S. E. (2018) Microhomology-mediated end joining: good, bad and ugly. *Mutat. Res.* **809,** 81–87 CrossRef Medline
81. Sfeir, A., and Symington, L. S. (2015) Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? *Trends Biochem. Sci.* **40,** 701–714 CrossRef Medline
82. Richardson, S. R., Morell, S., and Faulkner, G. J. (2014) L1 retrotransposons and somatic mosaicism in the brain. *Annu. Rev. Genet.* **48,** 1–27 CrossRef Medline
83. Nelson, P. N., Hooley, P., Roden, D., Davari Ejtehadi, H., Rylance, P., Warren, P., Martin, J., Murray, P. G., and Molecular Immunology Research Group (2004) Human endogenous retroviruses: transposable elements with potential? *Clin. Exp. Immunol.* **138,** 1–9 CrossRef Medline
84. Esnault, C., Maestre, J., and Heidmann, T. (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24,** 363–367 CrossRef Medline
85. Faulkner, G. J., and Garcia-Perez, J. L. (2017) L1 mosaicism in mammals: extent, effects, and evolution. *Trends Genet.* **33,** 802–816 CrossRef Medline
86. Ding, W., Lin, L., Chen, B., and Dai, J. (2006) L1 elements, processed pseudogenes and retrogenes in mammalian genomes. *IUBMB Life* **58,** 677–685 CrossRef Medline
87. Evrony, G. D., Cai, X., Lee, E., Hills, L. B., Elhosary, P. C., Lehmann, H. S., Parker, J. J., Atabay, K. D., Gilmore, E. C., Poduri, A., Park, P. J., and Walsh, C. A. (2012) Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **151,** 483–496 CrossRef Medline
88. Terry, D. M., and Devine, S. E. (2019) Aberrantly high levels of somatic LINE-1 expression and retrotransposition in human neurological disorders. *Front. Genet.* **10,** 1244 Medline CrossRef
89. Kapitonov, V. V., and Jurka, J. (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* **3,** e181 CrossRef Medline
90. Kapitonov, V. V., and Koonin, E. V. (2015) Evolution of the RAG1–RAG2 locus: both proteins came from the same transposon. *Biol. Direct.* **10,** 20 CrossRef Medline
91. Zhang, Y., Cheng, T. C., Huang, G., Lu, Q., Surleac, M. D., Mandell, J. D., Pontarotti, P., Petrescu, A. J., Xu, A., Xiong, Y., and Schatz, D. G. (2019) Transposon molecular domestication and the evolution of the RAG recombinase. *Nature* **569,** 79–84 CrossRef Medline
92. Liu, C., Yang, Y., and Schatz, D. G. (2019) Structures of a RAG-like transposase during cut-and-paste transposition. *Nature* **575,** 540–544 CrossRef Medline
93. Huang, S., Tao, X., Yuan, S., Zhang, Y., Li, P., Beilinson, H. A., Zhang, Y., Yu, W., Pontarotti, P., Escriva, H., Le Petillon, Y., Liu, X., Chen, S., Schatz, D. G., and Xu, A. (2016) Discovery of an active RAG transposon illuminates the origins of V(D)J recombination. *Cell* **166,** 102–114 CrossRef Medline
94. Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394,** 744–751 CrossRef Medline

95. Hiom, K., Melek, M., and Gellert, M. (1998) DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94,** 463–470 CrossRef Medline
96. Brandt, V. L., and Roth, D. B. (2004) V(D)J recombination: how to tame a transposase. *Immunol. Rev.* **200,** 249–260 CrossRef Medline
97. Lake, B. B., Ai, R., Kaeser, G. E., Salathia, N. S., Yung, Y. C., Liu, R., Wildberg, A., Gao, D., Fung, H. L., Chen, S., Vijayaraghavan, R., Wong, J., Chen, A., Sheng, X., Kaper, F., *et al.* (2016) Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352,** 1586–1590 CrossRef Medline
98. Treiber, C. D., and Waddell, S. (2019) Transposon expression in the *Drosophila* brain is driven by neighboring genes and diversifies the neural transcriptome. *bioRxiv* CrossRef
99. Leija-Salazar, M., Piette, C., and Proukakis, C. (2018) Somatic mutations in neurodegeneration. *Neuropathol. Appl. Neurobiol.* **44,** 267–285 CrossRef Medline
100. Maxwell, P. H. (2016) What might retrotransposons teach us about aging? *Curr. Genet.* **62,** 277–282 CrossRef Medline
101. Li, W., Prazak, L., Chatterjee, N., Grüninger, S., Krug, L., Theodorou, D., and Dubnau, J. (2013) Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nat. Neurosci.* **16,** 529–531 CrossRef Medline
102. Muotri, A. R., Zhao, C., Marchetto, M. C., and Gage, F. H. (2009) Environmental influence on L1 retrotransposons in the adult hippocampus. *Hippocampus* **19,** 1002–1007 CrossRef Medline
103. Dennis, S., Sheth, U., Feldman, J. L., English, K. A., and Priess, J. R. (2012) *C. elegans* germ cells show temperature and age-dependent expression of Cer1, a Gypsy/Ty3-related retrotransposon. *PLoS Pathog.* **8,** e1002591 CrossRef Medline
104. Van Meter, M., Kashyap, M., Rezazadeh, S., Geneva, A. J., Morello, T. D., Seluanov, A., and Gorbunova, V. (2014) SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat. Commun.* **5,** 5011 CrossRef Medline
105. De Cecco, M., Criscione, S. W., Peterson, A. L., Neretti, N., Sedivy, J. M., and Kreiling, J. A. (2013) Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging (Albany NY)* **5,** 867–883 CrossRef Medline
106. Kaeser, G. E., and Chun, J. (2020) Mosaic somatic gene recombination as a potentially unifying hypothesis for Alzheimer's disease. *Front. Genet.* **11,** 390 CrossRef Medline
107. Bekris, L. M., Yu, C. E., Bird, T. D., and Tsuang, D. W. (2010) Genetics of Alzheimer disease. *J. Geriatr. Psychiatry Neurol.* **23,** 213–227 CrossRef Medline
108. Brenowitz, W. D., Keene, C. D., Hawes, S. E., Hubbard, R. A., Longstreth, W. T., Jr., Woltjer, R. L., Crane, P. K., Larson, E. B., and Kukull, W. A. (2017) Alzheimer's disease neuropathologic change, Lewy body disease, and vascular brain injury in clinic- and community-based samples. *Neurobiol. Aging* **53,** 83–92 CrossRef Medline
109. Lee, M. H., and Chun, J. (2019) Mosaic APP gene recombination in Alzheimer's disease: what's next? *J. Exp. Neurosci.* **13,** 1179069519849669 CrossRef Medline
110. Chun, J. (2019) The gene conundrum in Alzheimer's disease. In *Cerebrum*, Dana Foundation, New York
111. Kern, D. M., Cepeda, M. S., Lovestone, S., and Seabrook, G. R. (2019) Aiding the discovery of new treatments for dementia by uncovering unknown benefits of existing medications. *Alzheimers Dement. (NY)* **5,** 862–870 CrossRef Medline
112. Herbert, A., and Rich, A. (1999) RNA processing in evolution: the logic of soft-wired genomes. *Ann. N.Y. Acad. Sci.* **870,** 119–132 CrossRef Medline
113. Herbert, A., and Rich, A. (1999) RNA processing and the evolution of eukaryotes. *Nat. Genet.* **21,** 265–269 CrossRef Medline
114. Davis, H. P., and Squire, L. R. (1984) Protein synthesis and memory: a review. *Psychol. Bull.* **96,** 518–559 CrossRef Medline
115. Crick, F. (1984) Memory and molecular turnover. *Nature* **312,** 101 CrossRef Medline