Brain cell somatic gene recombination and its phylogenetic foundations

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 within 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 aneuploidy and aneuposomies (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 genome-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 aneuposomic 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, hampere 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
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 intragenic junctions (IEJs), in which noncontiguous exons are fused by intragenic 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
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 in-frame gencDNAs APP-751 and 3/16 (in-frame examples include Australian, London, French, Indiana, and A673V; shown in red with yellow stars). 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
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 splicing 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 germ-line-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
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, genomic DNA 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 genomic DNA 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 genomic DNA 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 investigation. The potential age-related increases in SGR and genomic DNA 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.

---

### Table 1

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

<table>
<thead>
<tr>
<th>Mechanistic requirements</th>
<th>SGR</th>
<th>VDJ recombination</th>
<th>Retrotransposition of germline processed pseudogenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA strand break; mechanism</td>
<td>Yes; unknown</td>
<td>Yes; RAGs</td>
<td>Yes; retro-elements</td>
</tr>
<tr>
<td>Transcription</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Reverse-transcriptase activity</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>RNA intermediate</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>DNA repair process; mechanism</td>
<td>Likely; unknown</td>
<td>Yes; NHEJ</td>
<td>Likely; multiple</td>
</tr>
<tr>
<td>cis-Signal sequence</td>
<td>Yes; RSS</td>
<td>No</td>
<td>Unknown</td>
</tr>
<tr>
<td>Cell cycle/cell division</td>
<td>No</td>
<td>Unknown</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observed characteristics</th>
<th>SGR</th>
<th>VDJ recombination</th>
<th>Retrotransposition of germline processed pseudogenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actively transcribed</td>
<td>Yes (in vitro)</td>
<td>Yes</td>
<td>Variable</td>
</tr>
<tr>
<td>Number of recombed loci/cell</td>
<td>13 observed thus far</td>
<td>1</td>
<td>NA*</td>
</tr>
<tr>
<td>Intraexonic functions</td>
<td>Yes</td>
<td>No</td>
<td>None reported</td>
</tr>
<tr>
<td>Postmitotic somatic insertion</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Functional effects</td>
<td>APP coding diversity, insertional effects, and cell survival</td>
<td>Ig and T-cell receptor diversity</td>
<td>By definition, inactive</td>
</tr>
<tr>
<td>Genomic location</td>
<td>Non-wildtype loci; diverse</td>
<td>Wildtype locus</td>
<td>Constant, heritable genomic locus; non-wildtype</td>
</tr>
</tbody>
</table>

---

*NA, not applicable.*
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 memorizes 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 genomic and energetic efficiencies, albeit at the expense of plasticity.

Acknowledgments—We thank Danielle Jones and Dr. Laura Wolzonz 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, intron-exon 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


