nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-018-0718-6

In the format provided by the authors and unedited.

Somatic *APP* gene recombination in Alzheimer's disease and normal neurons

Ming-Hsiang Lee[1], Benjamin Siddoway[1,3], Gwendolyn E. Kaeser[1,2,3], Igor Segota[1,3], Richard Rivera[1], William J. Romanow[1], Christine S. Liu[1,2], Chris Park[1,2], Grace Kennedy[1], Tao Long[1] & Jerold Chun[1]*

[1]Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA. [2]Biomedical Sciences Program, School of Medicine, University of California, San Diego, La Jolla, CA, USA. [3]These authors contributed equally: Benjamin Siddoway, Gwendolyn E. Kaeser, Igor Segota. *e-mail: jchun@sbpdiscovery.org

Supplementary Information Discussion

Human neuronal *APP* gene recombination was identified in the normal and SAD brain, a conclusion supported by 9 distinct approaches (Extended Data Table 1). Recombination was characterized by the mosaic presence of thousands of distinct, genomic cDNAs (**gencDNAs**) that appeared as full and especially partial length *APP* sequences, containing precise exon::exon junctions, intra-exonic junctions (**IEJs**), insertions, deletions, and single nucleotide variations (**SNVs**). gencDNAs could be reproduced in culture via a process that required human *APP* RNA, reverse transcription and DNA strand breaks. At least some gencDNAs could be *in vitro* translated, producing toxic proteins and supporting functionality of the novel, recombined variants. gencDNAs showed both neuronal and SAD enrichment with up to 13 discernible loci in individual neurons that generally were distinct in location from wild type *APP* genomic loci. This enrichment occurred in 100% of a proof-of-concept SAD cohort and was accompanied by distinct variants including 11 SNVs that are considered pathogenic in FAD, and which were absent from non-diseased controls.

Relationship to prior studies

Neuronal gene recombination has not been previously reported. Human single-neuron genomic sequencing could theoretically reveal recombined loci[1-4] but is currently limited for CNV detection to ~1 Mb and more often several Mb vs. the 0.0007 Mb average size of gencDNAs[5]. Detection of gencDNAs using any single-neuron sequencing technique currently available[5] is further complicated by artifacts associated with each amplification method, varied bioinformatic approaches, low single-neuron sequencing success rates and very small sample sizes, and it is therefore not surprising that gencDNAs were not previously reported. For example, the highest

resolution approach capable of reporting gencDNAs by elegant single-cell nuclear transfer[4] was limited to mouse, had success rates of 1-2%, and sequenced just 7 total neurons from only the olfactory bulb of 4 mice, which could thus miss gencDNAs based on sample size and multiple biological differences including species, assessed neuroanatomical regions, ages, neuronal subtypes, and disease states. By contrast, alternative techniques as utilized in this study were required to detect recombination, especially the use of a diseased state – SAD – acting on a defined gene – *APP* – a general strategy that had been essential for identifying antibody gene rearrangement from a plasmacytoma neoplasm[6].

APP gencDNAs bear some resemblance to, yet are fundamentally distinct from, 1) processed pseudogenes[7] and 2) LINE1 repeat elements that themselves can encode an active reverse transcriptase (ORF2),[7] which along with human endogenous retroviruses (**HERVs**), telomerases, and possibly other RT sources could participate in gencDNA production. Processed pseudogenes were originally defined as evolutionary, non-coding, germline remnants of transcriptionally processed RNA[8]. However, some may be actively transcribed in cancer cells[9]. LINE1 elements are thought to have origins in ancient retroviruses and constitute over 17% of the human genome with an estimated 516,000 copies; however, only a very small subset is thought to be capable of retrotransposition in mitotic tissues[7,9] including neural progenitor cells (**NPCs**) of the developing brain[10]. The actual rate and biological significance of *bona fide* retrotransposition in human NPCs is a topic of active debate[2,10,11]. By comparison, and contrasting in multiple ways, neuron-enriched *APP* gencDNAs identified here manifest as thousands of distinct genomic variants marked by IEJs and myriad SNVs, undergo "retro-insertion" into post-mitotic neuronal genomes and can be actively transcribed and translated to produce bioactive products with relevance to both normal

and disease states. We note that some of the same molecular machinery, such as reverse transcriptase(s) could be shared amongst processed pseudogenes, LINE1 elements, and gencDNAs, albeit accessed at different developmental stages and in varied cell types.

gencDNAs arise through RNA intermediates, reverse transcription and DNA breaks

The genomic presence of gencDNAs identical to brain-specific RNA splice forms (APP 751 and 695 lacking exon 8 or 7 and 8, respectively) indicates that gencDNAs arise from spliced RNA intermediates generated in the brain. J20 mouse analyses further supported transcriptional RNA intermediates, in view of the neuron-specific transgene – present in the germline and thus all J20 cells – that displayed a striking age-dependent increase in gencDNAs only in neurons. Endogenous reverse transcriptase activity in CHO cells was required for gencDNA production based on the ability of nucleoside-based RT inhibitors to attenuate gencDNA formation in cell lines. Prevention of variant RNAs by reverse transcriptase inhibitors further supports their generation via gencDNAs. The cell type specificity and universality of this mechanism remain to be determined. Additional and/or different factors may be required to produce various forms of gencDNAs, in view of the distinct IEJ variant sequences detected in culture compared to human neurons. Further support for reverse transcription included high numbers of SNVs (Fig. 3), which are consistent with high polymerase error rates of the best studied reverse transcriptase in humans from HIV that exhibits 1/100th the fidelity of replicative DNA polymerases[12]. This could therefore account for the highly mutagenized *APP* gencDNAs observed which could be produced by successive rounds of reverse transcription within a single neuron. DNA breaks were also required, at least in cell culture, and in this light, the normal[13-15] and SAD brain[16,17] show myriad DNA breaks in adult neurons that might provide points of genomic integration for gencDNAs, including breaks produced by

neuronal activity-dependent processes[14,16]. The actual identity of endogenous reverse transcriptase(s) is currently unknown but might involve one or more forms of ORF2[7] within the 516,000+ LINE1 germline sequences, an issue that remains for future work.

gencDNA formation of *APP* variants bears striking parallels to RNA-mediated DNA recombination in mitotic yeast, where non-homologous and homologous recombination events are integrally related to genome rearrangements and gene expression[18]. The thousands of distinct gencDNAs thus far identified by SMRT sequencing combined with a vast majority of gencDNA DISH signals scattered throughout nuclei and distinct from WT loci (Fig. 2 o, p) support non-homologous recombination events, whereas results from J20 neurons showing an enlarging locus suggests the possibility of homologous events, both of which are not mutually exclusive. gencDNAs could also produce additional genome rearrangements distinct from the *APP* gencDNAs themselves. Entry of gencDNAs into the genome could be referred to as "retro-insertion" of gencDNAs within post-mitotic neurons, to reflect involvement of an RNA intermediate while remaining agnostic on mechanistic details. The concordance between gencDNA and mRNA sequences supports transcription from at least some retro-inserted gencDNAs, consistent with pseudogene transcription in yeast and human[8,18]. We do not know whether an initial gencDNA variant is first formed in RNA (through unknown mechanisms) followed by retro-insertion, then gencDNA transcription, or whether the observed full-length *APP* 695 and 751 gencDNAs (lacking IEJs) represent examples of precursors to other *APP* gencDNAs containing IEJs formed subsequently in DNA by micro-homology (or other mechanisms) followed by transcription. Importantly, these possible processes are not mutually exclusive, with both capable of producing concordant gencDNA and RNA sequences as we identified. In addition,

retro-insertion is compatible with processes that produce genomic mosaicism including genomic recombination, rearrangement and amplification mechanisms that are supported by rare FAD cases[19-21], but in SAD occur somatically and mosaically.

Relevance to SAD

Constitutive *APP* CNVs or mutations are considered causal in relevant FAD and Down syndrome cases, raising compelling possibilities for *APP* recombination mechanisms in SAD, where *APP* exonic CNVs were previously identified[22]; in light of the current data, these appear to reflect a minute part of a large universe of mosaic *APP* exon variation produced by somatic recombination. Our proof-of-concept data identified a dramatic shift in the forms and abundance of gencDNAs in SAD vs. non-diseased controls (Fig. 3 and 5), including the 3 to 5-fold increase in gencDNAs in all SAD brains examined (Fig. 5) with single neurons having up to 13 gencDNAs that were never observed in non-diseased brains. As remarkable was the identification of 11 different somatic SNVs previously published as pathogenic in FAD, which were absent from non-diseased controls. Other SNVs, as well as myriad genomic alterations identified in our data may also contribute to SAD and could function in both classical and non-classical mechanisms.

Classical mechanisms supporting the amyloid hypothesis involve production of toxic Aβ peptides and plaque formation. gencDNAs, including those with FAD mutations, likely represent a novel source of secretase-cleaved substrate to produce Aβ, as well as a potential source of toxic products not requiring secretase cleavage, as supported by cell culture studies (Extended Data Fig. 6). It is also possible that non-ATG translation could occur for out-of-frame variants produced by gencDNA integration near repeat elements, as reported in other neurological diseases[23]. Non-

classical mechanisms might involve RNA pathologies[24] or limits to gencDNA integration per neuron, beyond which neurodegenerative cell death might occur via genome instability produced by gencDNAs, akin to deleterious mobile elements[25]. Both diverse protein variants and non-protein mechanisms involving gencDNAs may help to explain the lack of efficacy observed for the many therapeutic trials targeting Aβ and related enzymologies, especially those targeting single molecular entities[26].

SAD increases with age, and the age-related increase of gencDNA variants in neurons compared to non-neuronal populations in J20 mice offers a possible explanation for the decades of life required for SAD to manifest. Neuronal *APP* transcription promotes gencDNA generation in both cell culture and in J20 neurons *in vivo* (Fig. 4 and 5), which is consistent with increased *APP* transcription previously linked to SAD incidence[27], and further mechanistically consistent with transcriptional augmentation of *APP* by the well-known SAD risk factor gene, ApoE4[28]. Intriguingly, gencDNA dependence on reverse transcriptase activities might be relevant to the statistical rarity of proven AD cases in HIV+ patients ≥65 years of age[29,30], who have received prolonged, combined anti-retroviral therapy (**cART**) that includes reverse transcriptase inhibitors. If confirmed, this observation would suggest immediate utility of the many FDA-approved cARTs or modified combinations thereof to treat SAD, DS and perhaps FAD. Additionally, processes producing DNA breaks such as head injury that have been linked to AD[16,31,32] are consistent with gencDNA production requiring DNA breaks. Taken together, gencDNAs and their production have properties relevant to a range of AD mechanisms and development of potentially novel therapeutics.

Implications for neuronal functions

The presence of some *APP* gencDNAs in non-diseased neurons likely reflects normal functions mediated by *APP*, which covers a gamut of possibilities[33,34] including synaptic function where *APP* gencDNAs might provide an increased repertoire of protein species, contributing to synaptic functional diversity. Additional genes may be transcriptionally modified and genomically re-encoded in response to selective activities in neuronal populations. Such a mechanism might enable preferential gene re-expression that bypasses splicing or further RNA modification. More broadly, gencDNAs could provide neurons with an activity-dependent mechanism to record and retain new genomic information over long periods of time, perhaps placing multiple forms of a gene under transcriptional control distinct from a wild type locus, which could be produced through diverse genomic integration sites. Such a process could have relevance to known neuronal functions that have already been shown to be dependent on transcriptional activity including Hebbian plasticity[35], synaptic wiring[36], memory[14] and cognitive function[37]. Thus, gencDNAs and their production may represent both a "recording" as well as a modifiable "playback" mechanism for expressing a "symphony" of gencDNA variants as well as wild type gene forms. It would be surprising if *APP* were the only gene to undergo this form of recombination, which altogether might impact distinct, normal brain functions as well as contribute to brain disorders, including AD.

1 Evrony, G. D. *et al*. Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49-59, doi:10.1016/j.neuron.2014.12.028 (2015).

2 Evrony, G. D., Lee, E., Park, P. J. & Walsh, C. A. Resolving rates of mutation in the brain using single-neuron genomics. *Elife* **5**, doi:10.7554/eLife.12966 (2016).

3 Lodato, M. A. *et al*. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94-98, doi:10.1126/science.aab1785 (2015).

4 Hazen, J. L. *et al*. The Complete Genome Sequences, Unique Mutational Spectra, and Developmental Potency of Adult Neurons Revealed by Cloning. *Neuron* **89**, 1223-1236, doi:10.1016/j.neuron.2016.02.004 (2016).
5 Rohrback, S., Siddoway, B., Liu, C. S. & Chun, J. Genomic mosaicism in the developing and adult brain. *Dev Neurobiol*, doi:10.1002/dneu.22626 (2018).
6 Hozumi, N. & Tonegawa, S. Evidence for somatic rearrangements of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. USA* **73**, 3628-3632 (1976).
7 Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* **24**, 363-367, doi:10.1038/74184 (2000).
8 Harrison, P. M., Zheng, D., Zhang, Z., Carriero, N. & Gerstein, M. Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res* **33**, 2374-2383, doi:10.1093/nar/gki531 (2005).
9 Kalyana-Sundaram, S. *et al*. Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* **149**, 1622-1634, doi:10.1016/j.cell.2012.04.041 (2012).
10 Muotri, A. R. *et al*. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**, 903-910, doi:nature03663 [pii] 10.1038/nature03663 (2005).
11 Upton, K. R. *et al*. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* **161**, 228-239, doi:10.1016/j.cell.2015.03.026 (2015).
12 Preston, B. D., Poiesz, B. J. & Loeb, L. A. Fidelity of HIV-1 reverse transcriptase. *Science* **242**, 1168-1171 (1988).
13 Blaschke, A. J., Staley, K. & Chun, J. Widespread programmed cell death in proliferative and postmitotic regions of the fetal cerebral cortex. *Development* **122**, 1165-1174 (1996).
14 Madabhushi, R. *et al*. Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell* **161**, 1592-1605, doi:10.1016/j.cell.2015.05.032 (2015).
15 Wei, P. C. *et al*. Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. *Cell* **164**, 644-655, doi:10.1016/j.cell.2015.12.039 (2016).
16 Suberbielle, E. *et al*. Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-beta. *Nat Neurosci* **16**, 613-621, doi:10.1038/nn.3356 (2013).
17 Mullaart, E., Boerrigter, M. E., Ravid, R., Swaab, D. F. & Vijg, J. Increased levels of DNA breaks in cerebral cortex of Alzheimer's disease patients. *Neurobiol Aging* **11**, 169-173 (1990).
18 Derr, L. K., Strathern, J. N. & Garfinkel, D. J. RNA-mediated recombination in S. cerevisiae. *Cell* **67**, 355-364 (1991).
19 McNaughton, D. *et al*. Duplication of amyloid precursor protein (APP), but not prion protein (PRNP) gene is a significant cause of early onset dementia in a large UK series. *Neurobiol Aging* **33**, 426 e413-421, doi:10.1016/j.neurobiolaging.2010.10.010 (2012).
20 Hooli, B. V. *et al*. Rare autosomal copy number variations in early-onset familial Alzheimer's disease. *Mol Psychiatry* **19**, 676-681, doi:10.1038/mp.2013.77 (2014).
21 Hooli, B. V. *et al*. Role of common and rare APP DNA sequence variants in Alzheimer disease. *Neurology* **78**, 1250-1257, doi:10.1212/WNL.0b013e3182515972 (2012).

22 Bushman, D. M. *et al*. Genomic mosaicism with increased amyloid precursor protein (APP) gene copy number in single neurons from sporadic Alzheimer's disease brains. *Elife* **4**, doi:10.7554/eLife.05116 (2015).
23 Cleary, J. D. & Ranum, L. P. Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders. *Curr Opin Genet Dev* **26**, 6-15, doi:10.1016/j.gde.2014.03.002 (2014).
24 Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243-247, doi:10.1038/nature22386 (2017).
25 McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792-801 (1984).
26 Egan, M. F. *et al*. Randomized Trial of Verubecestat for Mild-to-Moderate Alzheimer's Disease. *N Engl J Med* **378**, 1691-1703, doi:10.1056/NEJMoa1706441 (2018).
27 Brouwers, N. *et al*. Genetic risk and transcriptional variability of amyloid precursor protein in Alzheimer's disease. *Brain* **129**, 2984-2991, doi:10.1093/brain/awl212 (2006).
28 Huang, Y. A., Zhou, B., Wernig, M. & Sudhof, T. C. ApoE2, ApoE3, and ApoE4 Differentially Stimulate APP Transcription and Abeta Secretion. *Cell* **168**, 427-441 e421, doi:10.1016/j.cell.2016.12.044 (2017).
29 Turner, R. S. *et al*. An individual with human immunodeficiency virus, dementia, and central nervous system amyloid deposition. *Alzheimers Dement (Amst)* **4**, 1-5, doi:10.1016/j.dadm.2016.03.009 (2016).
30 Prevention, C. f. D. C. a. HIV Surveillance Report, 2016. 1-125 (Centers for Disease Control and Prevention, 2017).
31 Mortimer, J. A., French, L. R., Hutton, J. T. & Schuman, L. M. Head injury as a risk factor for Alzheimer's disease. *Neurology* **35**, 264-267 (1985).
32 Pravdenkova, S. V., Basnakian, A. G., James, S. J. & Andersen, B. J. DNA fragmentation and nuclear endonuclease activity in rat brain after severe closed head injury. *Brain Res* **729**, 151-155 (1996).
33 Nhan, H. S., Chiang, K. & Koo, E. H. The multifaceted nature of amyloid precursor protein and its proteolytic fragments: friends and foes. *Acta neuropathologica* **129**, 1-19, doi:10.1007/s00401-014-1347-2 (2015).
34 Nikolaev, A., McLaughlin, T., O'Leary, D. D. & Tessier-Lavigne, M. APP binds DR6 to trigger axon pruning and neuron death via distinct caspases. *Nature* **457**, 981-989, doi:10.1038/nature07767 (2009).
35 Guzman-Karlsson, M. C., Meadows, J. P., Gavin, C. F., Hablitz, J. J. & Sweatt, J. D. Transcriptional and epigenetic regulation of Hebbian and non-Hebbian plasticity. *Neuropharmacology* **80**, 3-17, doi:10.1016/j.neuropharm.2014.01.001 (2014).
36 Hattori, D., Millard, S. S., Wojtowicz, W. M. & Zipursky, S. L. Dscam-mediated cell recognition regulates neural circuit formation. *Annu Rev Cell Dev Biol* **24**, 597-620, doi:10.1146/annurev.cellbio.24.110707.175250 (2008).
37 West, A. E. & Greenberg, M. E. Neuronal activity-regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb Perspect Biol* **3**, doi:10.1101/cshperspect.a005744 (2011).

Supplementary Information Figure 1

Fig. 1b

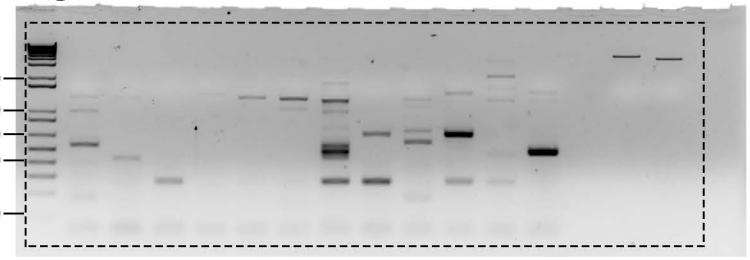
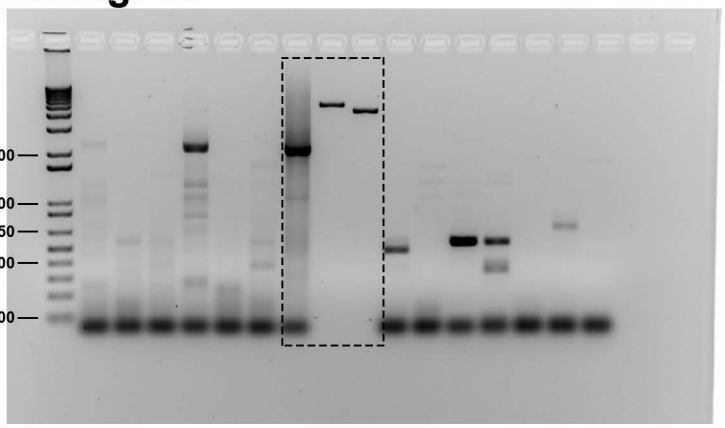
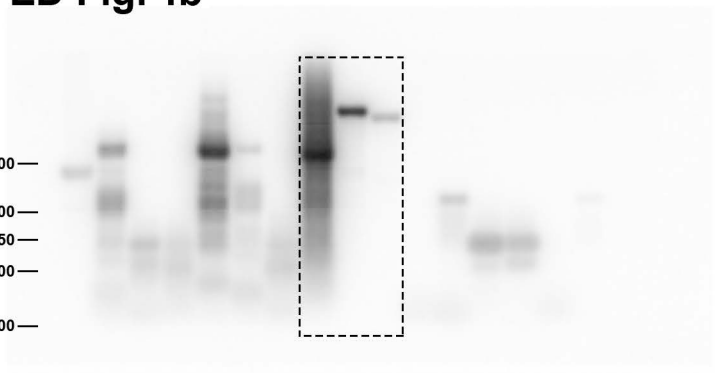


Fig. 1c

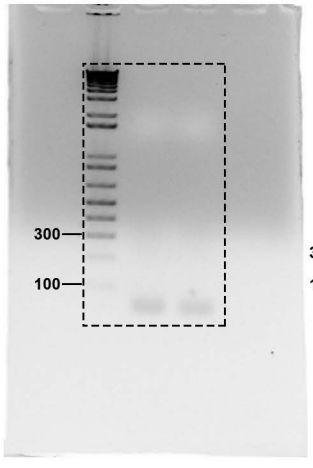
ED Fig. 1b



ED Fig. 1b



ED Fig. 1c



ED Fig. 1c

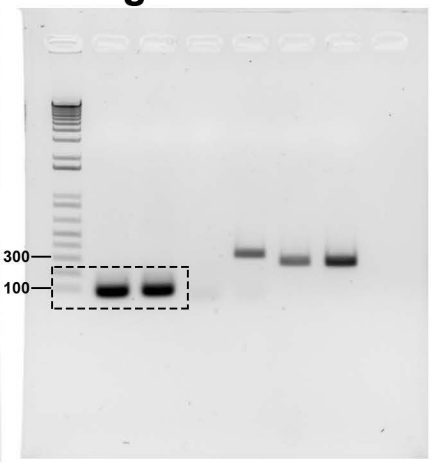


Fig. 2b, ED Fig. 2a

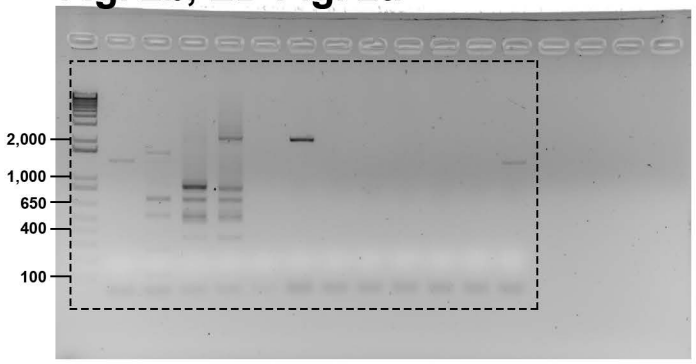
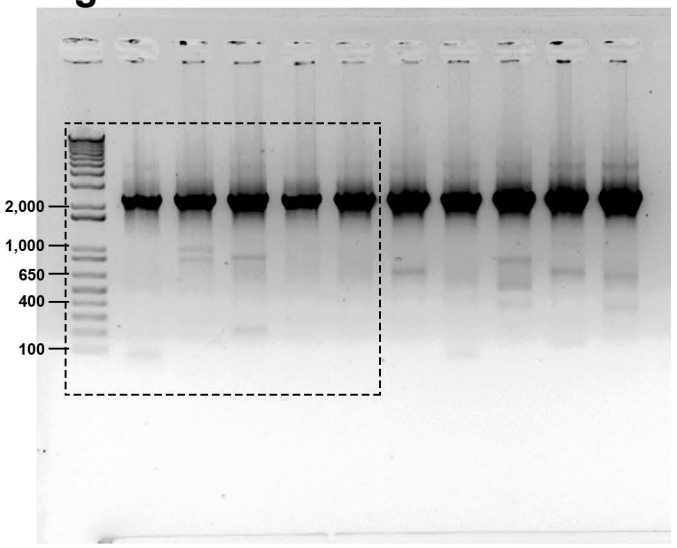
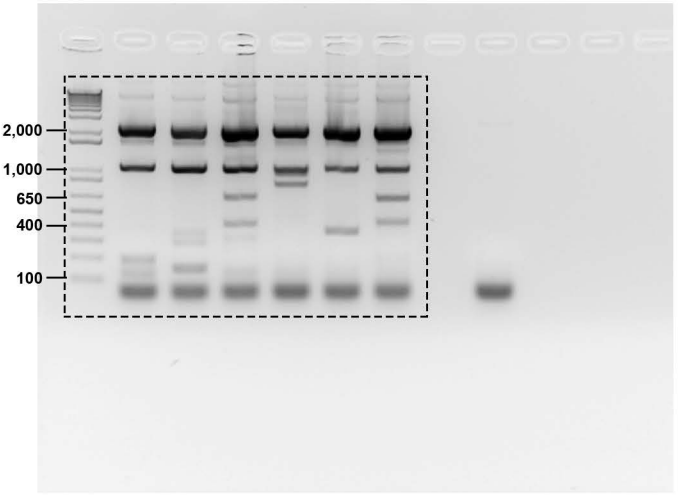


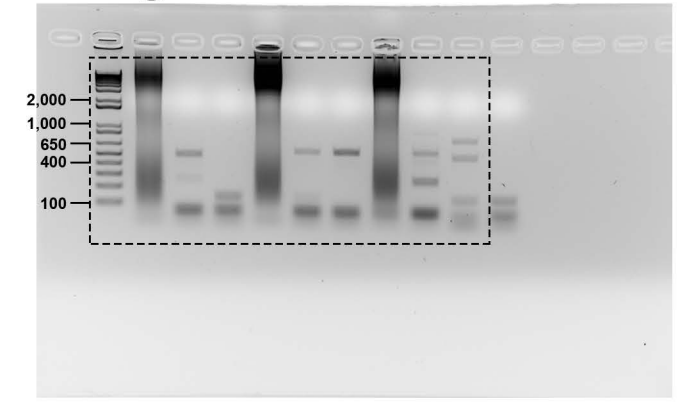
Fig. 4b



ED Fig. 1a

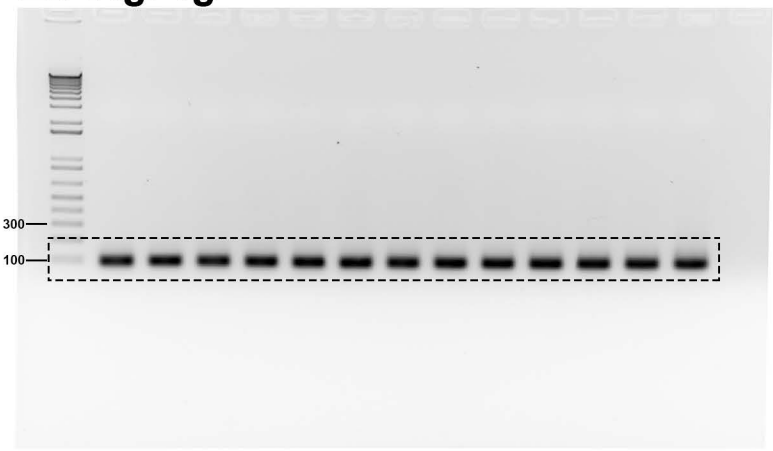


ED Fig. 1d

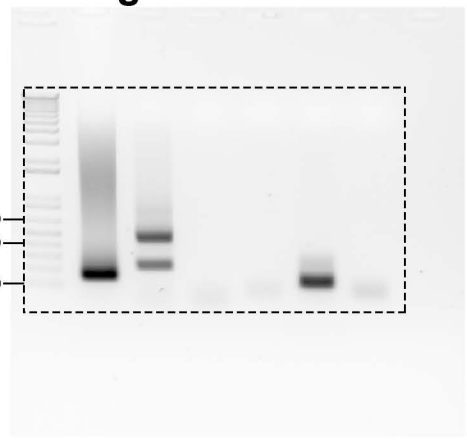


ED Fig. 1g

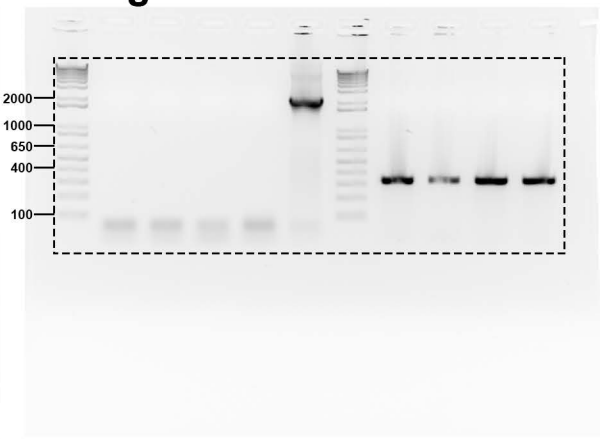
ED Fig. 1g



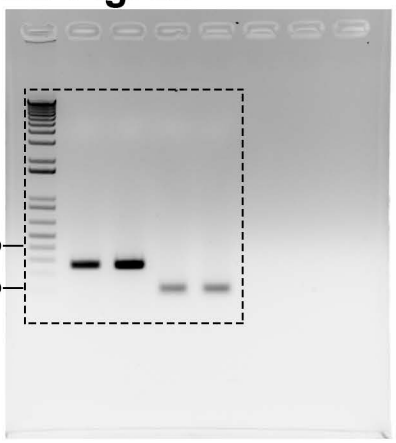
ED Fig. 2b



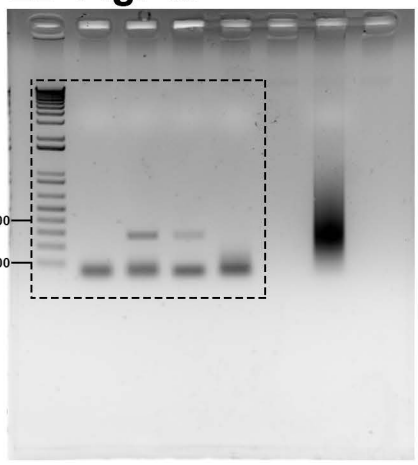
ED Fig. 2d



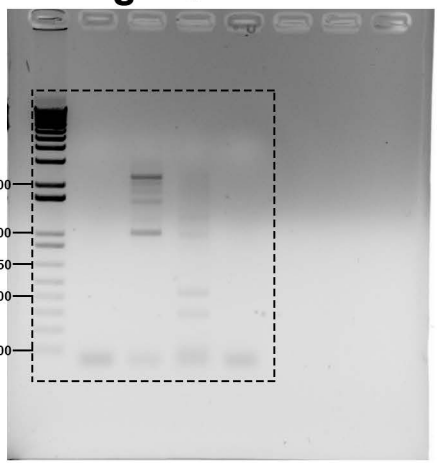
ED Fig. 2e



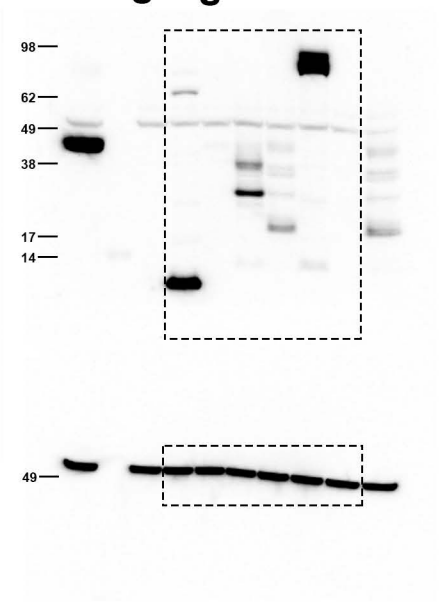
ED Fig. 4b



ED Fig. 4c



ED Fig. 5g



ED Fig. 4b, 4c