

Supplementary Methods

Agilent Human All Exon V6 pull-down experimental protocol and analysis

Human brain tissue

IRB-approved fresh frozen human brain tissue was provided by the University of California San Diego (UCSD) Alzheimer's Disease Research Center (ADRC) and the University of California Irvine (UCI) Institute for Mind Impairments and Neurological Disorders (MIND).

Nuclei extraction and fluorescence-activated nuclear sorting (FANS)

Human brain nuclei were isolated as described previously¹. Prior to human brain nuclei sorting, nuclei were labeled with anti-NeuN rabbit monoclonal antibody (1:800) (Millipore, Germany) and Alexa Fluor 488 donkey anti-rabbit IgG (1:500) (Life Technologies, Carlsbad, CA), and counterstained with propidium iodide (50 μ g/ml) (Sigma, St. Louis, MO). RNase A (100 μ g/ml) was included in all subsequent steps after initial nuclei isolation, including primary and secondary antibody incubations. Diploid NeuN-positive and negative nuclei were gated by PI and Alexa Fluor 488 fluorescence and sorted into appropriate populations for all human exon pull-down experiments. FANS was performed on a FACSAria Fusion (BD Biosciences, Franklin Lakes, NJ).

DNA extraction

DNA was extracted from isolated neuronal nuclei populations and precipitated with isopropanol. Briefly, nuclei were incubated with proteinase K in 550 μ l PK buffer (50mM Tris pH8.0, 0.1M EDTA, 0.1M NaCl, 1% SDS) overnight at 55°C. Samples were then treated with an RNase cocktail enzyme mix (ThermoFisher, Waltham, MA) for 2 hr, followed by the addition of 250 μ l

saturated NaCl. After centrifugation, DNA was precipitated from the supernatant with isopropanol and washed 3 times with 70% ethanol. Purified DNA was stored at -20 °C for future use.

Agilent SureSelect hybridization enrichment and sequencing

Nuclei were isolated from human prefrontal cortex via FANS. Genomic DNA was extracted and fragmented using sonication (Covaris, Woburn, MA). End repair reactions were performed and Illumina sequencing adaptors were ligated to genomic DNA. Library-prepped DNA was hybridized with Agilent SureSelect probes designed against all human exons (V6 without UTRs). Pulled-down genomic DNA sequences were then sequenced on an Illumina NextSeq (Illumina, San Diego, CA).

Identification of *APP* geneDNAs and insertion sites from 10 pull-down datasets

Fastq files from the Park *et al.*² datasets were downloaded from SRA (accession PRJNA532465). Sequences from the Park *et al.* and Chun Lab datasets were first aligned to the human reference genome (GRCh38) using BWA-mem (version 0.7.17) and screened with Vecuum (version 1.0.1) to confirm that APP plasmid was not detected in any of these datasets. Datasets were then aligned to the human genome using STAR (version 2.7.3a) with the options previously described in Lee *et al.*³ (--outSAMattributes All --outSJfilterCountTotalMin 1 1 1 1). We excluded reads that contained the tag 'jI:Bi:i,1' using SAMtools (version 1.3) in order to keep the reads where a junction was detected. To remove reads that were “mis-split” by STAR, we filtered out reads that had fewer than 16bp mapped to each side of the junction (at least 16bp of each exon are required to improve mapping accuracy). Duplicate reads were marked and removed using Picard (version

2.1.1). Reads were then processed and visualized using a modified version of the R exonjunction package (<https://github.com/christine-liu/exonjunction>).

To identify APP novel insertion sites, we first extracted discordant read pairs in which one read mapped to an APP exon::exon junction or was clipped (≥ 20 bps) at APP UTR ends, and the mate read mapped confidently (mismatches $< 1\%$) to a novel chromosome other than chromosome 21. We also made sure that these reads mapped uniquely to the human genome but not to the mouse genome or transcriptome. Next, we mapped the UTR clipped sequences against the human genome using BLAST+ (version 2.9.0) and kept only the reads confidently aligned (query coverage $\geq 90\%$, identical matches $\geq 98\%$) to the same chromosomes as their corresponding mates. Detailed read alignment information can be found in Supplementary Table 1.

Supplementary Discussion of gencDNAs in blood

The source of blood gencDNAs is not known but could be from lymphocytes that can express all *APP* forms⁴, or could suggest that neural gencDNAs can be detected in blood and may have value as a potential biomarker.

Supplementary Discussion of DISH and RISH

DISH and RISH are important to our conclusions and cannot be ignored. Our DISH and RISH probes were custom designed with Advanced Cell Diagnostics (ACD, Newark, California) to target the exon::exon junction between exon 16 and 17 and the IEJ between exon 3 and 16. Multiple controls and experiments validated the specificity of our ISH probes (*Lee* Figure 2j-n, Extended Data Figure 1f, 2f-h, 6b, 6d-f)³: 1) sense and antisense probes that demonstrated DNA specificity of the sense-strand probe; 2) targeted restriction enzyme digestion that effectively destroyed the target locus and eliminated probe-dependent signal; 3) double-labeling that showed gencDNAs were usually distinct from the endogenous locus; 4) creation and use of retrovirally-delivered synthetic targets in cell culture that confirmed probe specificity; and 5) use of the J20 mouse model that contains the human *APP* sequence that confirmed specificity by the sense-strand identification of the transgene. ACD and others have demonstrated the ability of BaseScope™ probes to differentiate exon::exon junctions⁵, splice variants, and point mutations⁶, and therefore, we are confident that our RISH and DISH probes identify *APP* gencDNAs and IEJs in a PCR-free and plasmid-free system. Additionally, DISH resulted in cell type and disease-specific patterns of gencDNA expression, with SAD neurons showing the greatest number of gencDNA foci, while J20 mouse experiments demonstrated specificity as shown in *Lee et al.* None of these results could be explained by PCR of *APP* plasmid sequences.

Supplementary References

- 1 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).
- 2 Park, J. S. *et al.* Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat Commun* **10**, 3090, doi:10.1038/s41467-019-11000-7 (2019).
- 3 Lee, M. H. *et al.* Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **563**, 639-645, doi:10.1038/s41586-018-0718-6 (2018).
- 4 Schlossmacher, M. G. *et al.* Detection of distinct isoform patterns of the beta-amyloid precursor protein in human platelets and lymphocytes. *Neurobiol Aging* **13**, 421-434 (1992).
- 5 Visualize EGFRvIIIsplice variant in glioblastoma sample: Enabled by BaseScope Assay <https://acdbio.com/science/applications/research-areas/egfrviii>. (2019).
- 6 Baker, A. M. *et al.* Robust RNA-based in situ mutation detection delineates colorectal cancer subclonal evolution. *Nat Commun* **8**, 1998, doi:10.1038/s41467-017-02295-5 (2017).

ReadName	Read (with APP)	Mate (novel site)	Read orientation
SRR8898224.96962287	CGTGCTGGTGCATCTTGGAG	GCCTCATGTTGCTCATTGGT	R1R2
SRR8898224.107824814	CTTCCCCGTCATCATCGGCT	GAGTGGGATGCCACATGTC	R1R2
SRR8898224.20867379	CTCCACCTCAGCCTCTTCTI	TAAAATTGAAGTACAGGCTA	F1R2
SRR8898224.118878929	CCTTGATCACTAATCATGT	TGAGAGGAAAAGGCCCTT	F1R2
SRR8898276.166141196	CGCTCCTCTGGGGTGACAG	CAACATCCAATCATGTATAA	F1R2
SRR8898224.107643979_mate	GCGAGTGGGATCCGCCGCC	GAAGGCCCTGGAAACCGAC	F2R1
SRR8898224.27622836_mate	CCGTCCCCTGTCACTGGG	GGGGTGAAGGAGGTGCCA	F2R1
SRR8898224.49640101_mate	TTACCAATTCTGCCTCCTGTI	CCTGGGACTGCTCCTTTACT	F2R1
SRR8898224.96736610_mate	ACAATGTCTTTTAAATTAATA	CCTCTTCCTTCACACCACCC	F1R2
SRR8898268.233861544_mate	CAGCAACCAAGGCTCATACT	GCTCTGGAAACAGGAAGAA	F2R1

Read location on APP	Mate location	Sample ID	Region type	Intersected Gene
18/17	chr3:72631698-72631798	AD304 HIF	intergenic	intergenic
6/5	chr13:37663681-37663781	AD304 HIF	Intron	TRPC4
6/5	chr5:140870048-140870148	AD304 HIF	Intron	PCDHA1
14/13	chr1:155946754-155946854	AD304 HIF	3'UTR	ARHGEF2
18/17	chr6:72497946-72498046	AD315 blood	intergenic	intergenic
5'UTR	chr1:148953252-148953352	AD304 HIF	Intron	PDE4DIP
3'UTR	chr12:7210081-7210181	AD304 HIF	5'UTR	PEX5
3'UTR	chr3:172916391-172916491	AD304 HIF	Exon	SPATA16
3'UTR	chr9:112879022-112879122	AD304 HIF	3'UTR	SLC46A2
3'UTR	chr10:43845520-43845620	AD317 HIF	Intron	LINC00619