Figure S1

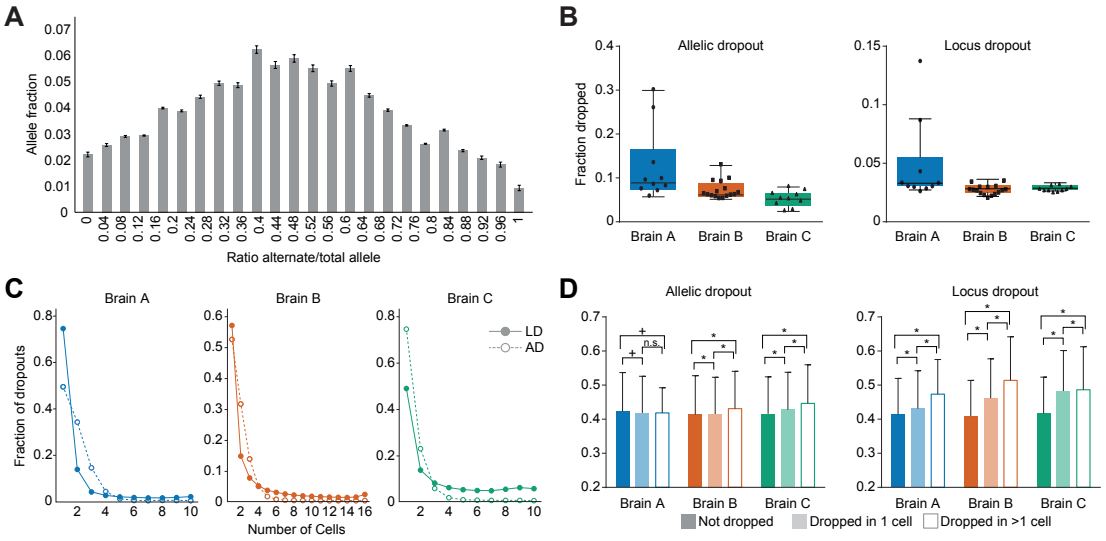

Figure S1. The genome is reliably amplified from single nuclei. (A) Analysis of allelic balance at confident heterozygous sites suggests that allele fractions vary, but are centered on 0.5. Loci at which one allele was completely dropped out are not shown. (B) Analysis of allelic (left) and locus (right) dropout rate in Brains A, B and C. Two outlier neurons in Brain A increase its median rates, while other Brain A neurons closely resemble those from Brains B and C. (C) Recurrence analysis of allelic dropout (AD; light colors) and locus dropout (LD; dark colors) across single cells. Approximately 50% of dropouts are unique to a single cell, while the rest recur in multiple cells. (D) Analysis of GC content of dropped loci. Recurrently dropped loci tend to have higher GC content then loci dropped in one cell, or not dropped, except for AD in Brain A, which shows a slight but significant bias in the opposite direction. * denotes an increased GC content in group with more dropouts, $p < 0.05$, 2-way ANOVA with Sidak's multiple testing correction. + denotes decrease GC content in group with more dropouts, $p < 0.05$, 2-way ANOVA with Sidak's multiple testing correction.

Figure S2

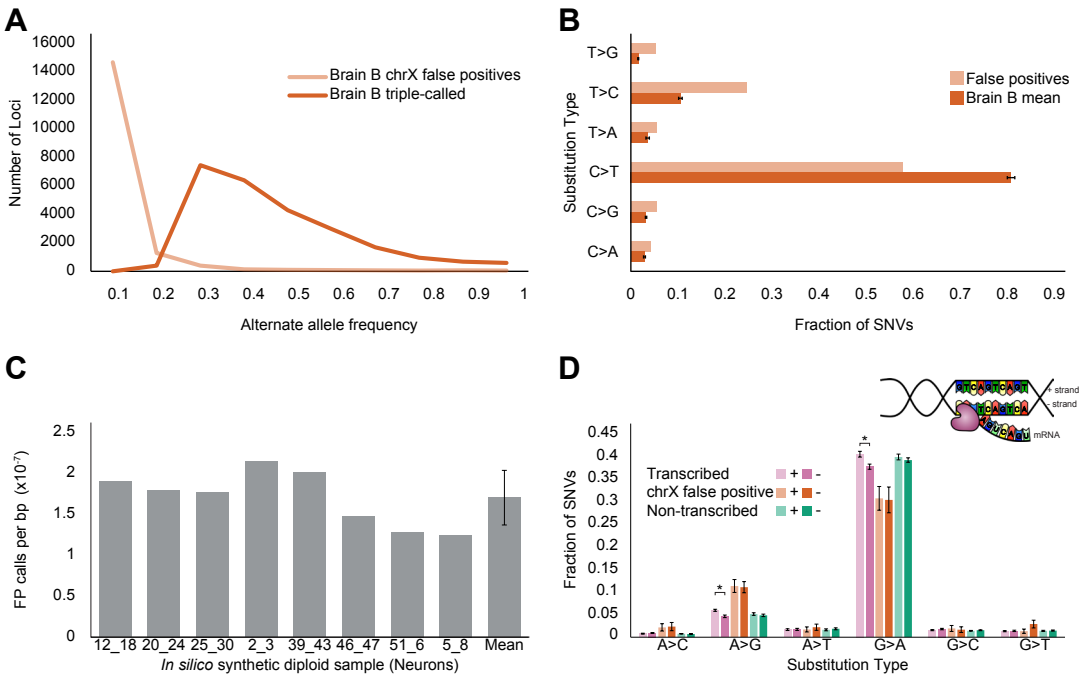

Figure S2. Triple-called SNVs and false positives in Brain B have distinct sequence characteristics. (A) Distribution of alternate allele fractions of X-chromosome false positives defined at the read level (see Methods) and candidates called by all three callers in Brain B. Alternate alleles in triple-called variants are present at higher fraction of total depth than false positives. The distribution of alternate allele fractions for triple calls is not centered on 0.5, but is skewed somewhat to the left. This may be partially due to our triple calls being comprised of ~23% false positive loci. Another factor may be single-stranded, unrepaired DNA lesions which existed in the neuronal genome at the time the cells were harvested. Such single-strand lesions would be expected to display an alternate allele fraction centered on 0.25. (B) The SNV substitution types that comprise false positives and triple calls is distinct, suggesting they occur by distinct mechanisms (technical vs. biological, respectively). Chi-square test with Yates correction p-value < 0.01. (C) The rate of triple-called X-chromosome false positives from in silico synthetic X-chromosome neurons. Samples are named for the X-chromosome data sets which were combined, so sample "12_18" is the combined X-chromosome data from Brain B Neurons 12 and 18. (D) Comparison of strand bias at transcribed candidates in Brain B, non-transcribed loci (+ and – strand of genomic DNA based on the hg19 build was used to substitute for template and non-template strands), and false positives in transcribed loci. Purine-purine transitions were enriched on the non-transcribed strand in somatic SNVs (*, $p < 0.05$ ANOVA with Sidak's correction for multiple testing) but not at FP loci, suggesting FPs were not generated by transcriptional damage, and that the strand bias phenomenon reflects a signature of true mutations. Non-transcribed loci display no bias, as expected.

Figure S3

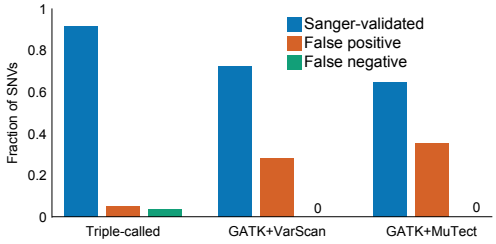

Figure S3. Validation of Brain B SNVs by Sanger sequencing. Validation rate of SNVs called by all three callers, those called by GATK and each of two somatic callers, respectively. False negatives are variants that were called homozygous reference in bulk tissue, but were found to be heterozygous in bulk upon Sanger validation.

Figure S4

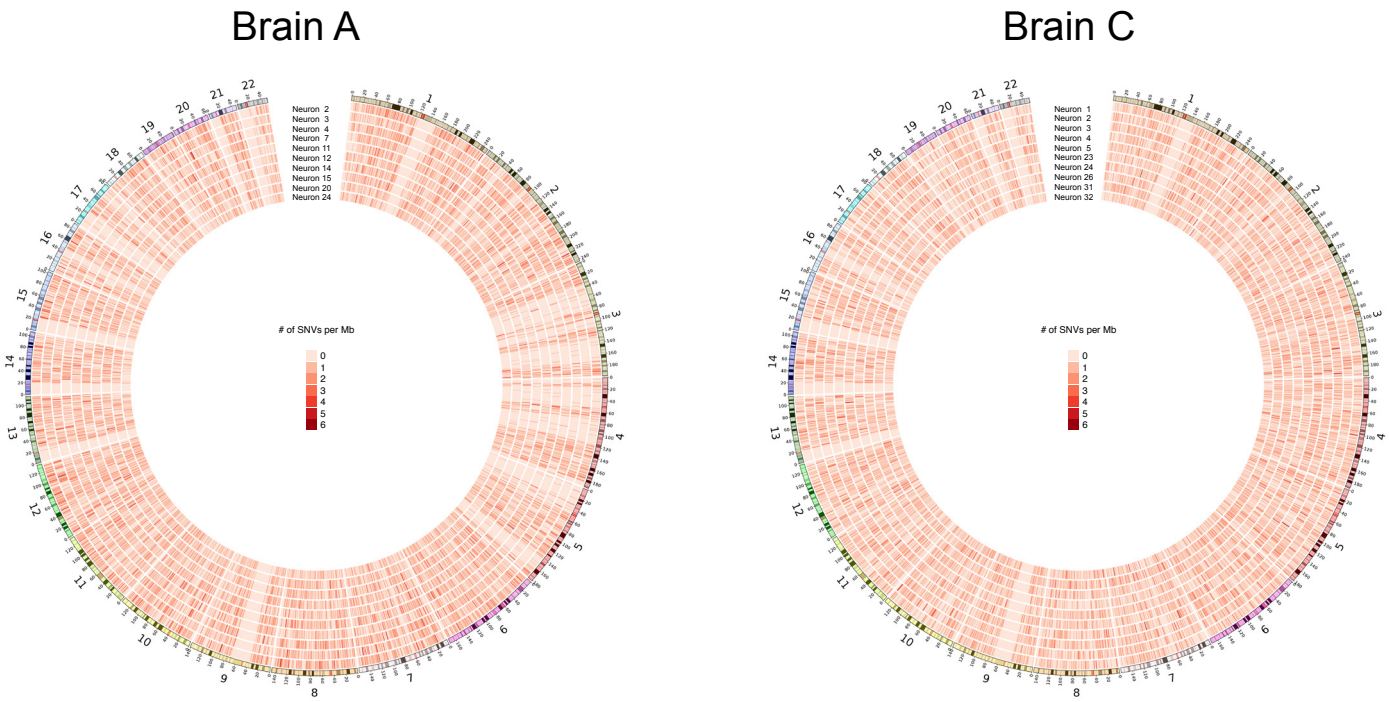

Figure S4. Circos plots of SNV rate per megabase in Brain A and Brain C across human autosomes. Somatic SNVs are distributed broadly across the genome in neurons derived from these samples.

Figure S5

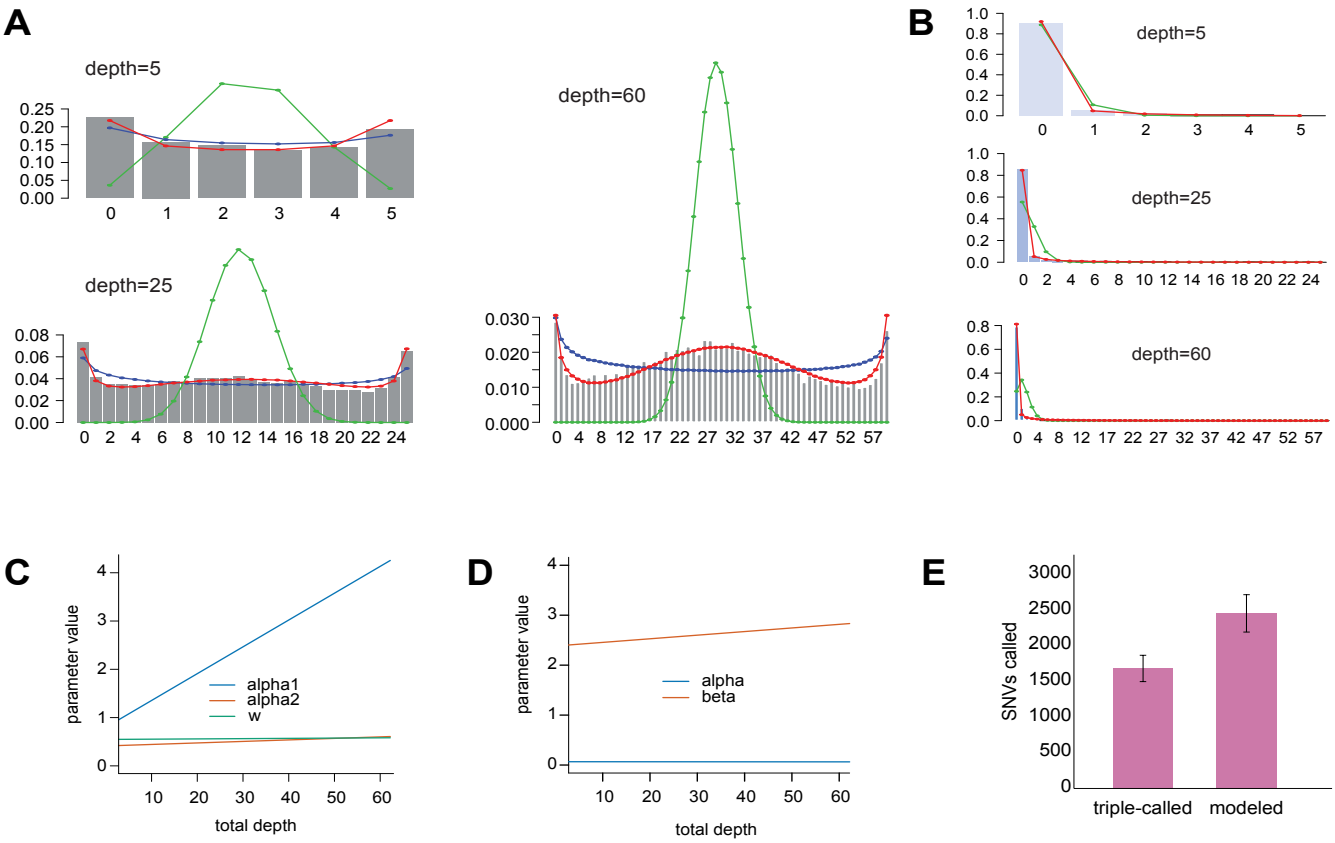

Figure S5. Single-cell MDA SNV modeling in Brain B. (A) Distributions of reads supporting the alternate allele, conditioned on the total read depth, for Brain B Neuron 8, for confident heterozygous loci, with curves representing the expected fraction from binomial (green), beta-binomial (blue) and our final beta-binomial mixture linear model (red). (B) Same as S5A for nonvariant homozygous loci, with curves representing the expected fractions from binomial (green) and our final beta-binomial linear model (red). (C, D) Parameters of the heterozygous (C) and nonvariant homozygous (D) models based on the estimated slopes and intercepts with respect to total read depth. (E) Comparison of number of SNVs called per neuron in Brain B by the two approaches. The triple-calling method gives 31% fewer SNVs compared to this model-based method.

Figure S6

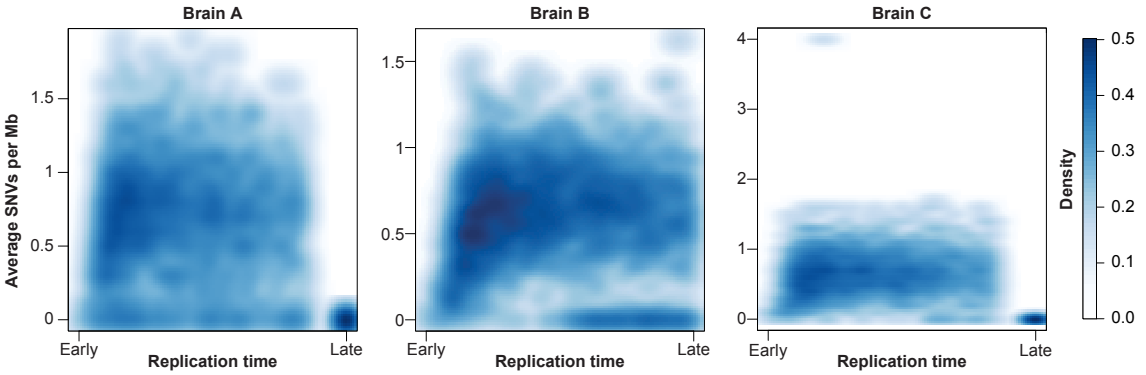

Figure S6. Single-neuron SNVs do not strongly correlate with replication timing in Brains A, B, or C. (Spearman correlation = -0.07, 0.1, and -0.02 for Brains A, B, and C, respectively). Density of blue points reflects density of SNVs at that position on the plot.

Figure S7

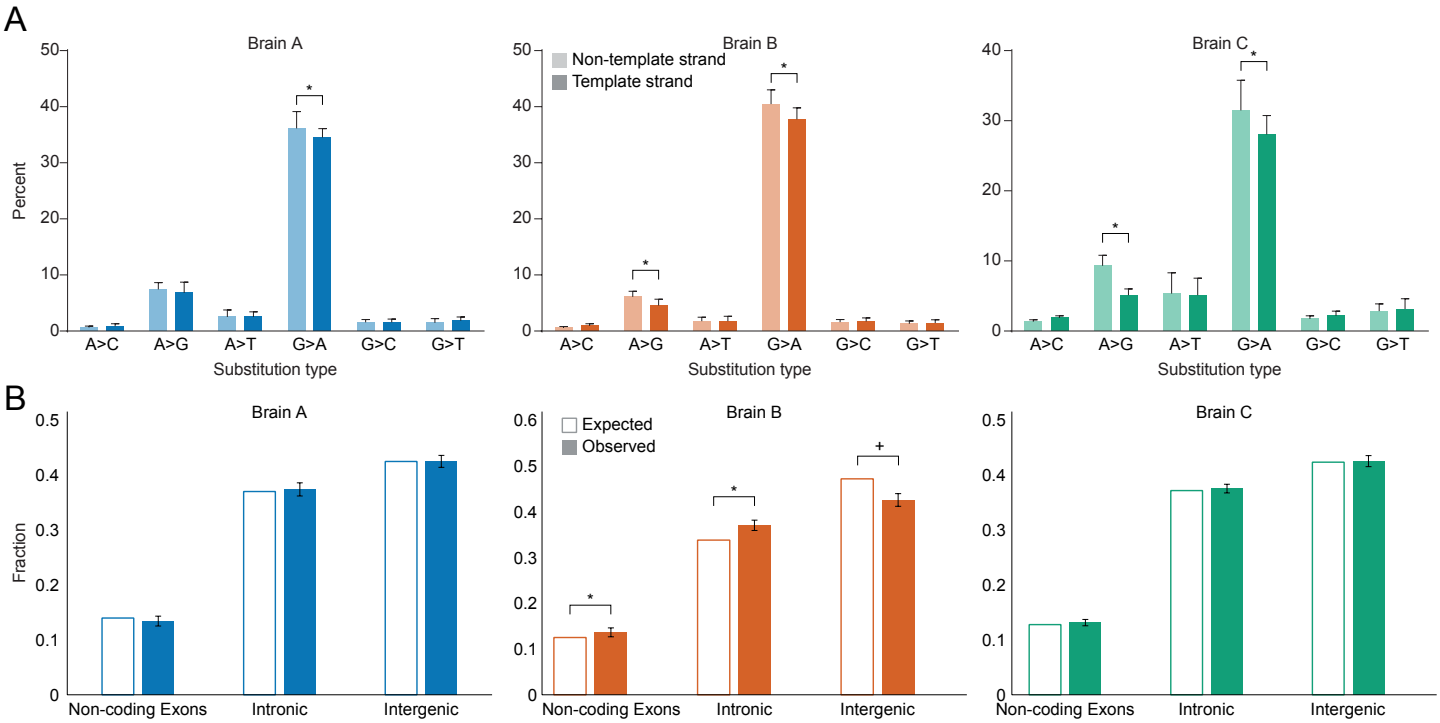

Figure S7. Analysis of transcription-associated SNVs (A) Analysis of strand bias for all substitution types in transcribed regions. Only purine-purine transitions display strand bias. * denotes $p < 0.05$, ANOVA with Sidak's correction for multiple testing (B) Fraction of SNVs in Brains A, B, and C in non-coding exons, intronic loci, and intergenic regions. *, $p < 0.05$, combined binomial distribution for enrichment of SNVs with Bonferroni correction, + $p < 0.05$, binomial distribution for depletion of SNVs with Bonferroni correction. Exons were significantly enriched for SNVs in all brains (see Figure 2), while introns in Brain A and C, and non-coding exons in Brain C, displayed non-significant enrichment of SNVs.

Figure S8

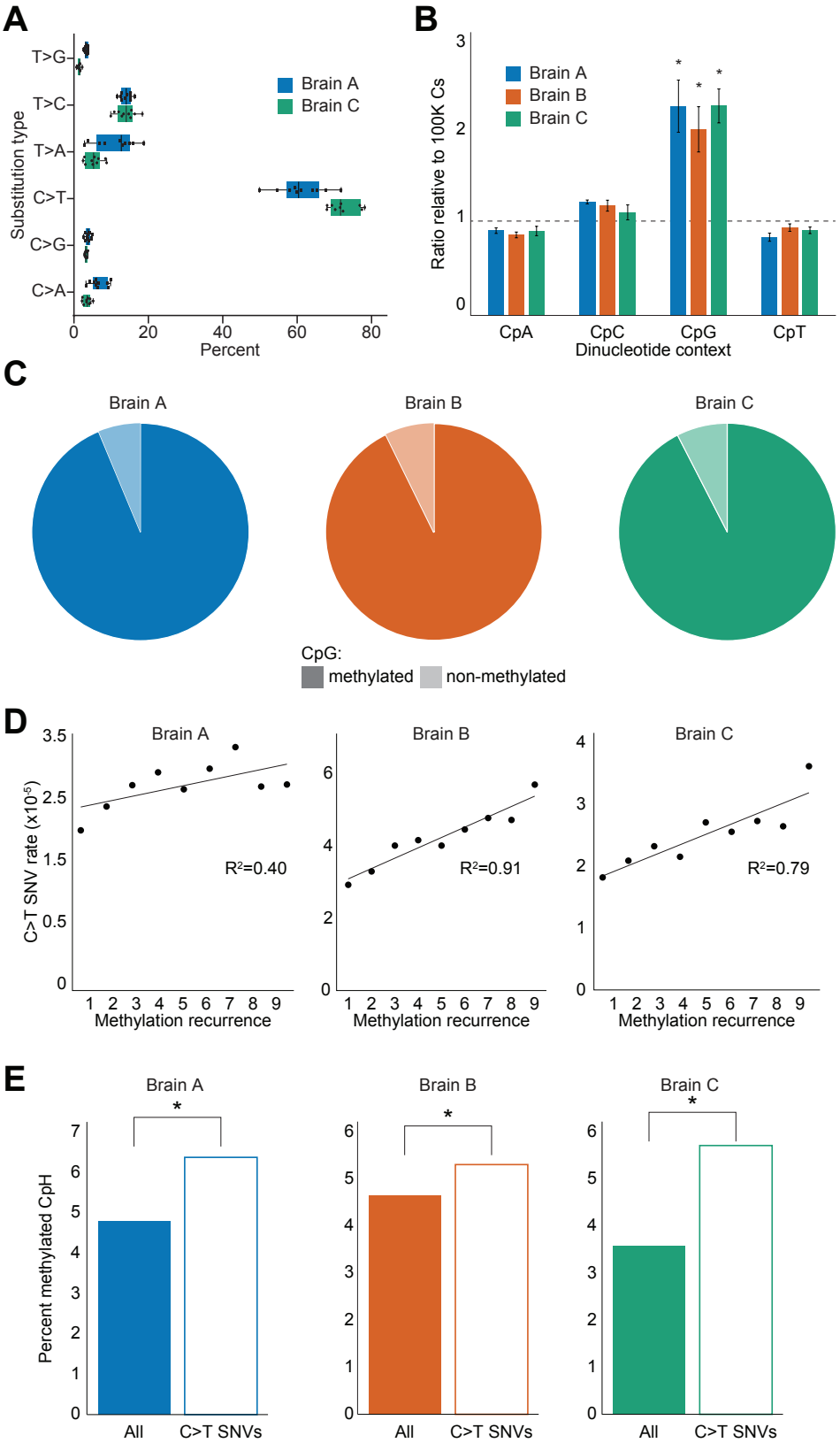

Figure S8. Relationship between single-neuron SNVs and cytosine methylation. (A) SNS spectrum of SNVs observed in Brain A and Brain C. Similar to Brain B, SNVs are biased toward C>T transitions. (B) Cytosines mutated to thymine in single neurons were approximately twice as likely to be located in a CpG context, in which most cytosines are methylated. (C) 92%-94% of somatic CpG>TpG transitions occurred at an meC (D) The CpG > TpG mutation rate of a given cytosine was correlated with the frequency with which that cytosine was methylated during human life.(E) Non-CpG methylation (meCpH) is also common in the brain, and single-neuron SNVs were enriched for meCpH . *, $p < 0.001$, Chi-Square with Yates correction. Thus, the deamination of methylated cytosine to thymidine significantly impacts the somatic genome of normal human neurons.

Figure S9

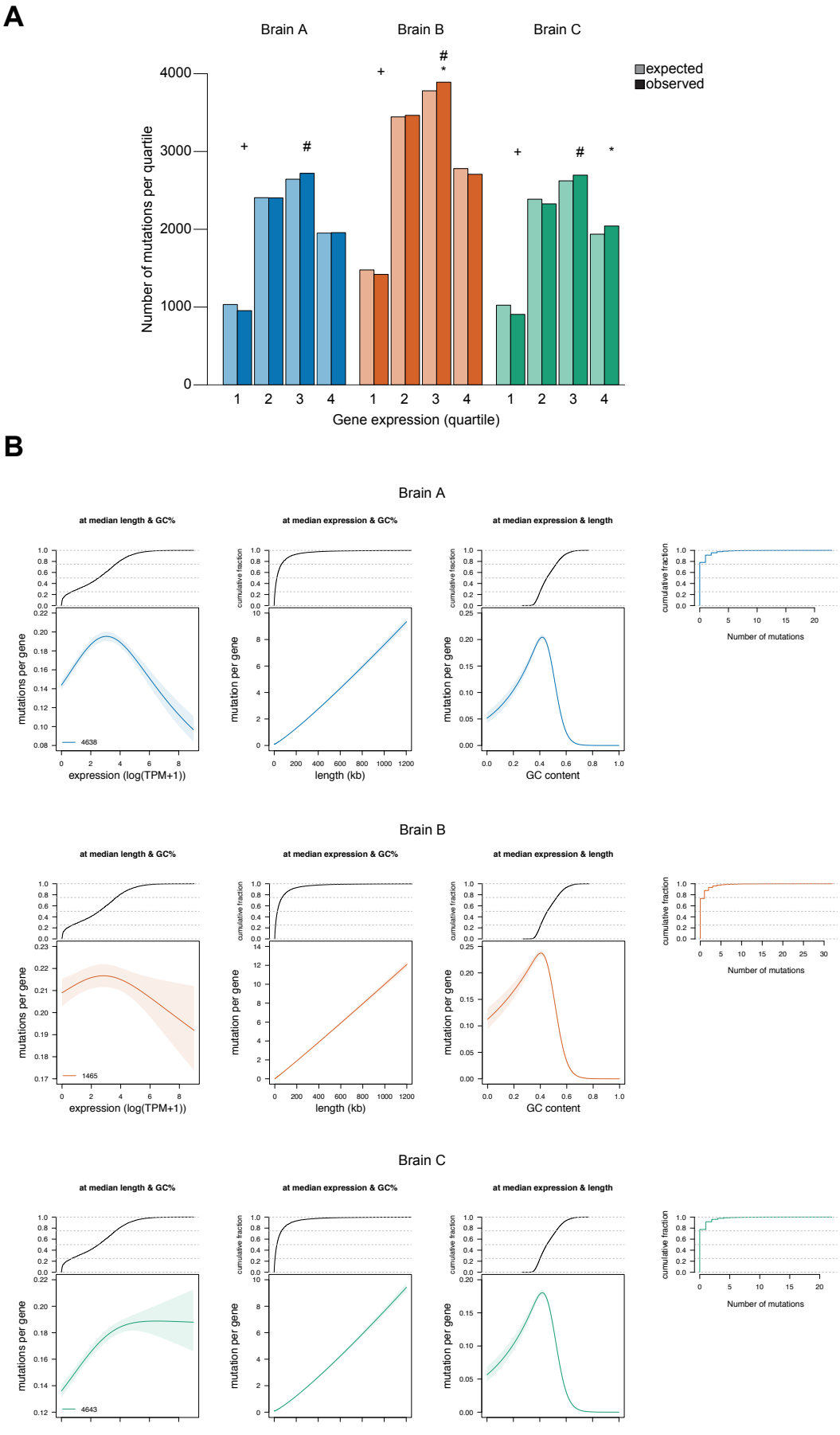
A
Brain A
Brain B
Brain C
expected
observed
Number of mutations per quartile
Gene expression (quartile)
B
Brain A
Brain B
Brain C
at median length & GC%
at median expression & GC%
at median expression & length
mutations per gene
mutation per gene
cumulative fraction
expression (log(TPM+1))
length (kb)
GC content
Number of mutations
4638
1465
4643

Figure S9 (cont'd)

Figure S9. Genes involved in neuronal processes are enriched for somatic SNVs. (A) Observed and expected number of SNVs in gene expression quartiles. + denotes significant depletion, Poisson test p-value < 0.05 for individual brains; * denotes significant enrichment, Poisson test p-value < 0.05 for individual brains. # denotes enrichment, p < 0.05, Fisher's combined p-value. (B) The effect of expression level, length, GC content on the number of mutations per gene, estimated using a generalized additive model. Each plot shows the effect of one variable fixing the other two at their medians. Expression level is shown in log (TPM+1). The relative shape of each function is invariable, though the absolute Y values depend on the other two variables. Shading represents the standard error. The number of mutations peaks near the third quartile of the expression levels, increases linearly with gene length and peaks around 40% GC. Plots in the rightmost column depict the number of SNVs observed per gene.

Figure S10

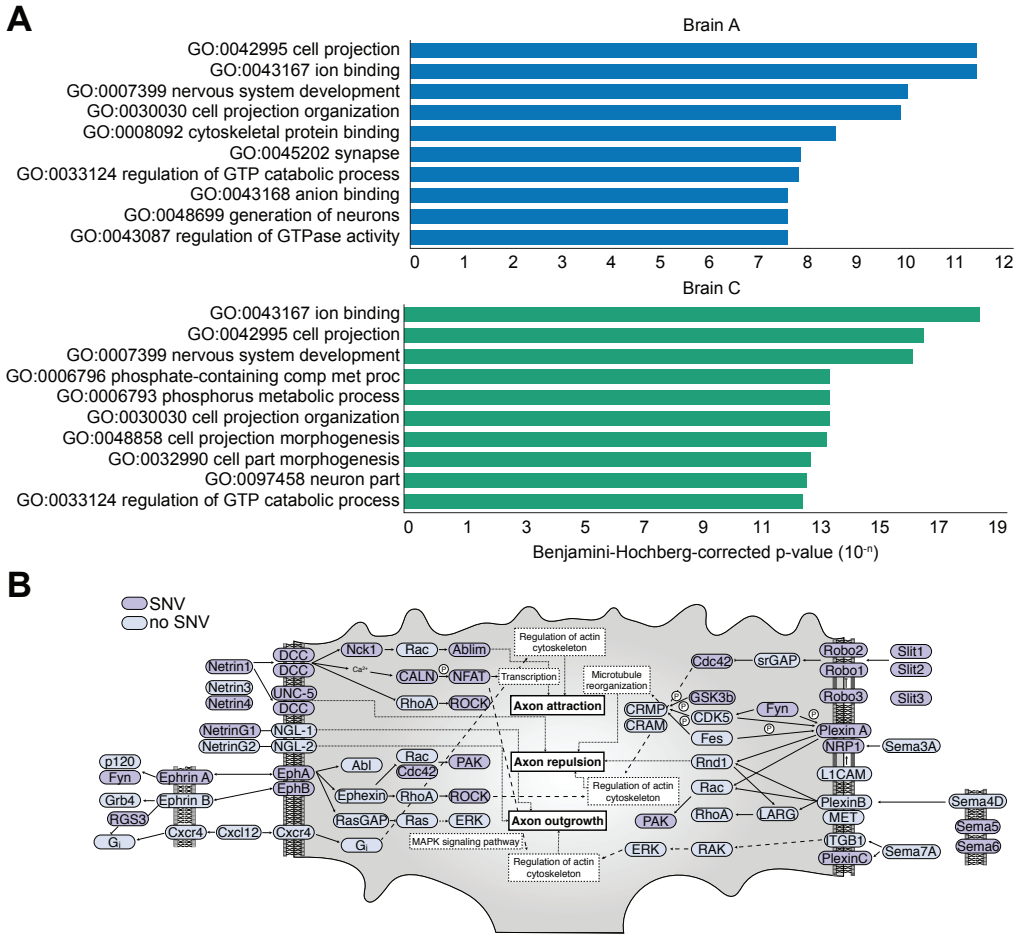

Figure S10. Genes involved in neural processes are enriched for single-neuron SNVs. (A) Gene ontology analysis of mutated genes in Brains A and C. (B) KEGG pathway analysis of genes in the “Axon Guidance” pathway mutated in Brain B neurons. Purple boxed genes were mutated in at least one neuron.

Figure S11

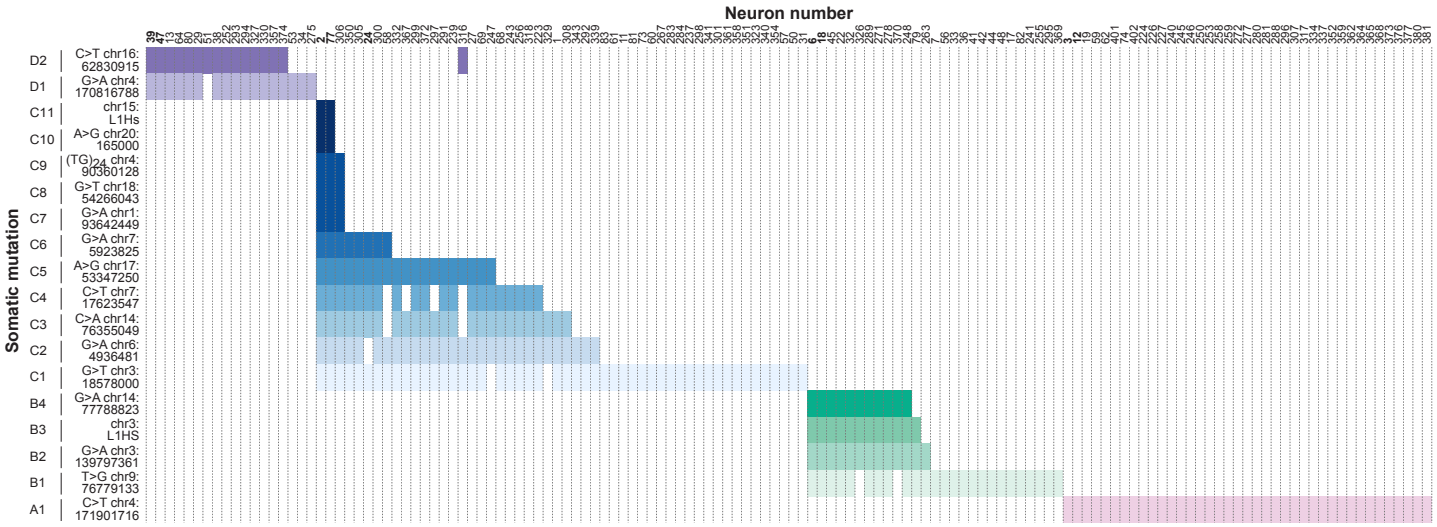

Figure S11. Single-cell genotyping of somatic mutations demonstrates single-neuron SNVs in Brain B are widespread and resemble nested clones. 136 neurons from Brain B positive for at least one of 18 somatic variants. Colored cells reflect presence of somatic variants, blank cells represent cells in which the variant was not detected. Blank cells within clades (for example, Neuron 58 at chr7:17623547 C>T, or Neuron 326 at chr9:76779133 T>G) would have been expected to be present based on other variants present in those cells. These mutation dropouts most likely represent amplification failure (see Figure S2), though other explanations are possible. One anomalous cell (Neuron 316) has mutations reflective of membership in both the purple and blue clades, and likely represents a rare doublet sorting error.

Figure S12

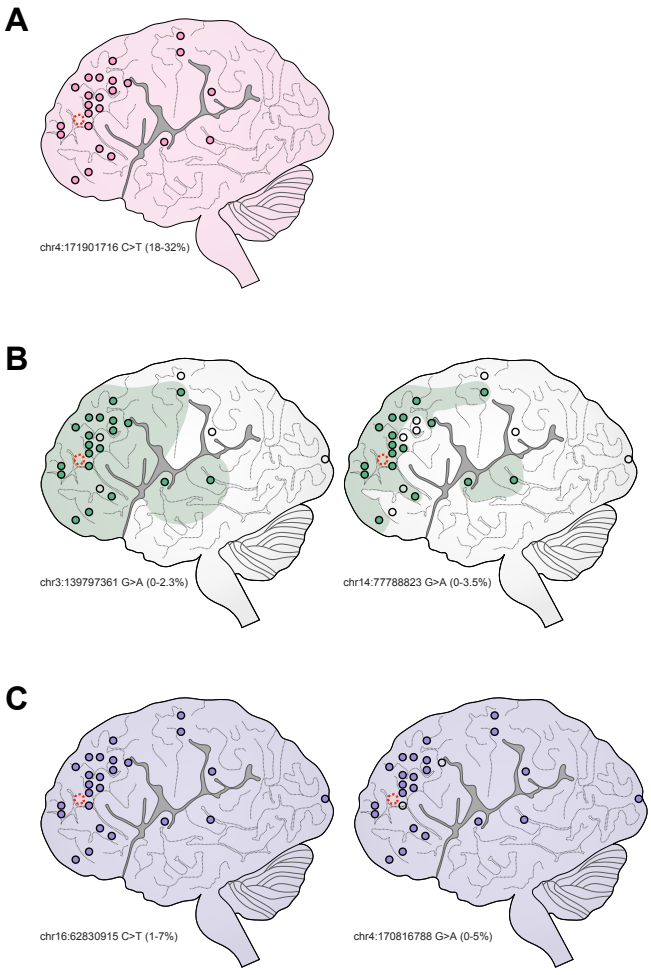

Figure S12. Distribution of shared SNVs across cortical areas by ultra-deep sequencing of 25 cortical samples. (A) chr4:171901716 C>T (pink) is distributed across all cortical areas at high mosaicism, varying from 18% to 32%. (B) chr3:139797361 G>A and chr14:77788823 G>A (green) are distributed widely across cortical areas, although they are present at low mosaicism (0-2.3% and 0-3.5%, respectively). (C) chr16:62830915 C>T and chr4:170816788 G>A are present at moderate mosaicism (0-7%) across all areas of the cortex, as well as other tissues in the body (Fig. 3C).

Figure S13

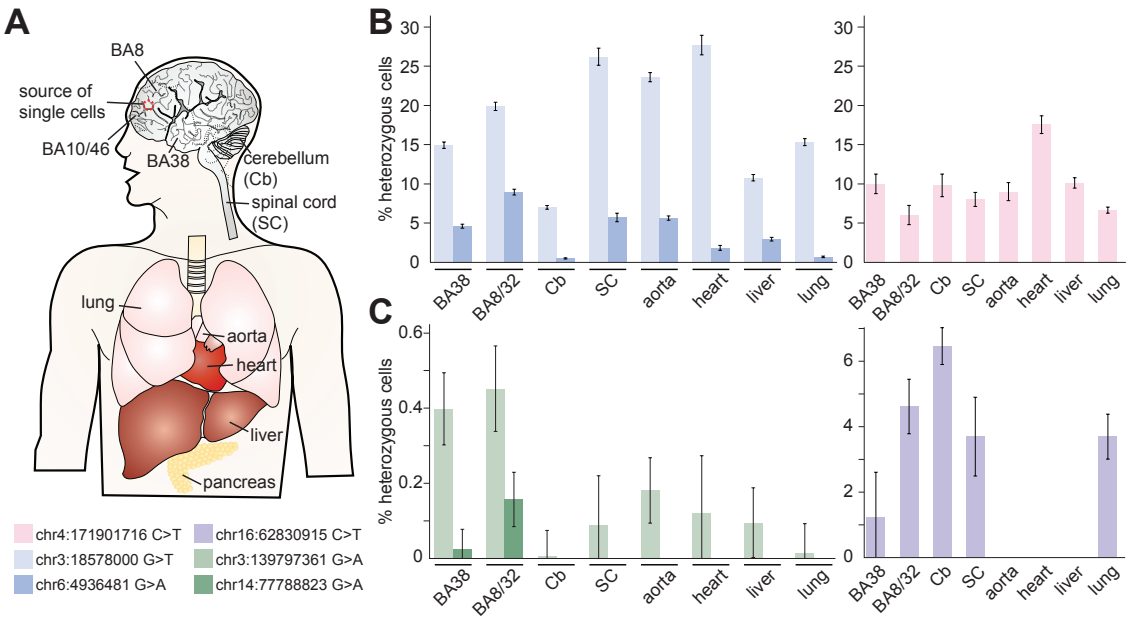

Figure S13. Allele-specific droplet digital PCR genotyping of selected SNVs from bulk unamplified DNA. (A) Schematic of brain and body areas genotyped by ultra-deep sequencing (Fig. 3C). (B) Droplet digital PCR analysis confirms the presence of Brain B shared somatic mutations in unamplified DNA, that variants present in a larger number of single neurons also show proportionately higher allele fractions as well as broader tissue distribution, and that mosaic variants present in only a fraction of neurons are distributed widely across the body, indicating that they arose very early in development. ddPCR analysis and ultra-deep sequencing (Fig. 3C) independently arrive at similar mosaic percentages.

Figure S14

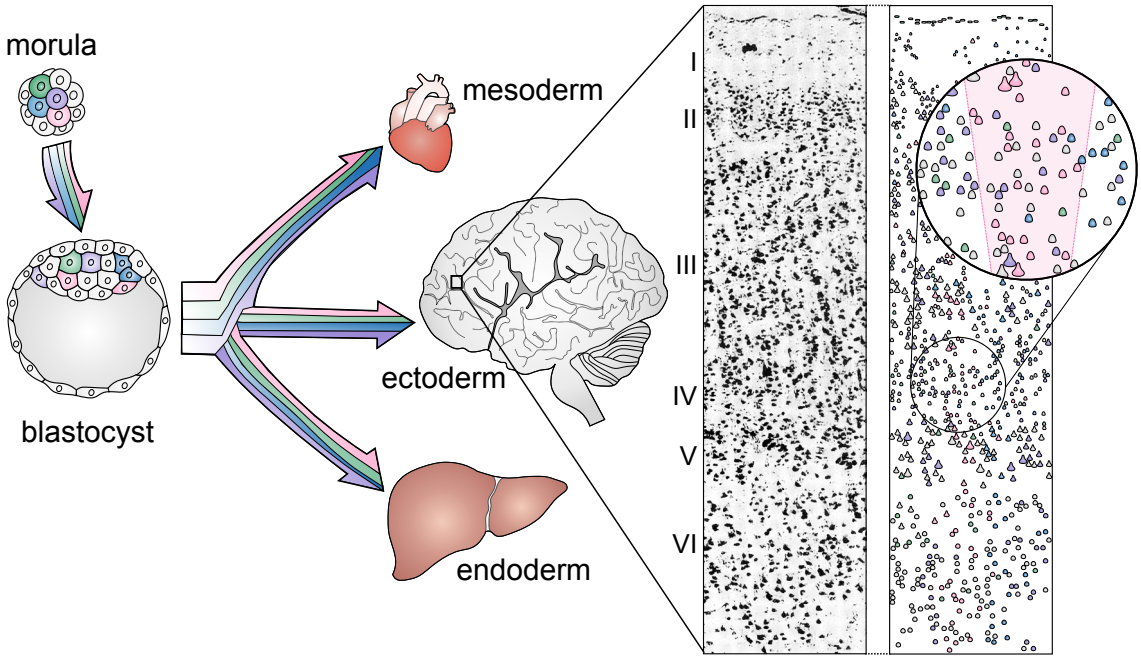

Figure S14. Model: Somatic mutations acquired early in embryogenesis result in the profoundly polyclonal architecture of the mature cortex. Cells that acquire somatic mutations very early in development, for example, blastomeres in the morula (pink, green, blue, or purple) form clones of pluripotent cells, which then segregate into organs derived from all three germ layers (left panel). Within the cerebral cortex, these pre-gastrulation somatic mutations mark spatially adjacent but deeply clonally divergent neurons (right panel). Although clonal dispersion may be restricted tangentially (for example, pink neurons bounded by pink background, inset), closely related neurons are extensively intermingled with neurons that share no clonal connection closer than gastrulation. Cresyl violet stain of human primary somatosensory cortex in right panel is modified from (56).