Somatic mutation in single human neurons tracks developmental and transcriptional history


Neurons live for decades in a postmitotic state, their genomes susceptible to DNA damage. Here we survey the landscape of somatic single-nucleotide variants (SNVs) in the human brain. We identified thousands of somatic SNVs by single-cell sequencing of 36 neurons from the cerebral cortex of three normal individuals. Unlike germline and cancer SNVs, which are often caused by errors in DNA replication, neuronal mutations appear to reflect damage during active transcription. Somatic mutations create nested lineage trees, allowing them to be dated relative to developmental landmarks and revealing a polyclonal architecture of the human cerebral cortex. Thus, somatic mutations in the brain represent a durable and ongoing record of neuronal life history, from development through postmitotic function.

To investigate rates and patterns of somatic SNVs, we analyzed high-coverage (-40x) whole-genome sequencing (WGS) data from 36 single neurons from postmortem brain tissue of three neurotypical individuals: a 15-year-old female (UMB4638), a 17-year-old male (UMB4645), and a 42-year-old female (UMB4643) (sources designated brains A, B, and C, respectively). Using fluorescence-activated nuclear sorting (FANS), we purified single NeuN cortical neuronal nuclei from the prefrontal cortex of postmortem human brain tissue; we then amplified the DNA by multiple-displacement amplification (MDA) (Fig. 1A). We subjected this amplified DNA to high-throughput sequencing, achieving $\geq 10x$ average coverage per neuron at 82 to 87% of loci (table S1). After analysis of overall quality (8), false-positive and false-negative variant calls, allele balance, and locus/allele dropout (fig. S1 and S2; supplementary text), we used these single-cell MDA WGS data to identify somatic mutations, focusing on SNVs.

We identified single-cell SNV candidates by means of three established mutation-calling algorithms (see materials and methods), thereby defining a conservative list of somatic SNVs as those identified by all three callers in at least one single neuron but absent from DNA isolated from bulk tissue (e.g., heart) from the same individual (Fig. 1B). These “triple-called” SNVs were confirmed by Sanger sequencing at a very high rate (92%) in the single-cell DNA sample from which they were identified (table S2 and fig. S3). Single neurons from the three brains averaged 1685 to 1793 triple-called SNVs (Table 1, Fig. 1D, table S3, and fig. S4). Across all neurons, we observed a mean of 8.3% allelic dropout (loss of one allele of a heterozygous locus) and 3.3% locus dropout (fig. S1). With an estimated 23% false-discovery rate, our results suggest that each neuron from these individuals may have contained 1458 to 1580 somatic SNVs. Additional model-based analyses resulted in a similar range of SNVs identified (fig. S5; supplementary text). These rates of mutation per neuron are consistent with single-cell sequencing of other normal cell types (22) but are lower than somatic SNV rates in normal skin cells, which are exposed to damaging ultraviolet light (13), and in several tumor types (11, 12, 14).

The molecular profile of neural SNVs is quite distinct from cancer and germline mutations, and reveals mutagenic forces that affect a neuron during its life. Cancer and germline mutation rates correlate with DNA replication (15, 16), with late-replicating DNA more susceptible to mutation, and they negatively correlate with transcription (15, 17). In contrast, single-neuron SNVs did not correlate with replication timing (fig. S6 and table S4) and were enriched in coding exons (Fig. 2A) with a strand bias (purine-purine transitions enriched on the nontranslating strand), as expected for transcription-associated mutations (18, 19) (Fig. 2B and fig. S7). We also observed a signature of methylated cytosine (meC) to thymine (T) transitions (fig. S8), which can occur as a result of replication-independent deamination of meC in single-neuron SNVs. Taken together, these data demonstrate that replication-independent mutational mechanisms generate more SNVs than does replication in human neurons, which are postmitotic and live long, transcriptionally active lives.

Integration of our SNV data with gene expression in the prefrontal cortex from public databases (20) revealed that highly expressed genes (in particular, those in the third quartiles) were enriched for single-neuron SNVs with statistical significance (Fig. 2D, fig. S9, and table S5), and, unlike mutations in cancers (such as glioblastoma multiforme), single-neuron SNVs correlated with chromatin markers of transcription (21) from fetal brain (Fig. 2C and table S6). Neural-related gene sets were enriched for somatic SNVs (Fig. 2E, fig. S10, and table S7), and single neurons harbored heterozygous coding SNVs in genes that, when a single copy is mutated in the germline, confer a high risk of neurological disease (table S8). For example, SCN1A (seizure disorder) (22) and SLC22A2 (schizophrenia) (23) both contained coding mutations in neurons from brain B (Fig. 2F and table S8). Thus, genes active in human neurons and critical for their function are vulnerable to somatic mutation, and even the normal brain contains individual neurons with disruptive mutations.

Mutations shared by multiple neurons, which must necessarily have arisen during development, revealed surprising lineage relationships in the human brain. We genotyped shared somatic variants in brain B, including 15 SNVs (table S9), a T/G dinucleotide expansion on chromosome 4 (see materials and methods), and two L1 retrotransposon insertions (7, 8), in 210 amplified single-neuron genomes isolated from the same region of cortex as the original 16 cells from brain B; 136 of 226 (60%) single-neuron genomes contained at least one clonal SNV (fig. S11), and the results suggested serial mutations over the course of development. For example, three SNVs showed...
distinct mosaic profiles: SNV C1 was present in 23% (51/226) of the neurons tested (suggesting that it occurred early in development), C8 was present only in three of these 51 neurons, and C10 only in two of this set of three, neurons 2 and 77 (Fig. 3A and fig. S11). SNVs identified mutually exclusive sets of neurons such that, for example, cells marked by variants discovered in neurons 39 and 47 never contained SNVs present in neurons 6 and 18, 2 and 77, or 3 and 12 (except one anomalous cell; fig. S11). These data placed 9/16 sequenced cells, and 60% (136/226) of analyzed cortical neurons, into four separate clades (pink, green, blue, and purple in Fig. 3A), which suggests that at least five distinct clades—the four marked, plus at least one additional unmarked clade—gave rise to the 226 neurons in this sample.

Genotyping DNA samples from across the brain, using ultradepth sequencing of a targeted panel of clonal somatic SNVs, showed that although some SNVs were present at low mosaic fractions (up to 4% of cells) only in restricted regions of frontal cortex, all four major clades dispersed across the cortex at surprisingly low levels of mosaicism (green clade, 0.2 to 2%; blue clade, 0.5 to 6%; purple clade, 1 to 7%; pink clade, 18 to 32%; Fig. 3B, fig. S12, and table S10). For example, SNVs C8 and C10 were restricted to <3% of cells in the middle frontal gyrus surrounding the site of sequencing, whereas C1 marked deeper branches of this same clade and was found throughout the cortex in 7 to 26% of cells, as well as in the cerebellum and spinal cord (Fig. 3B). Analysis of non-brain tissue of individual B showed that variants present in >5 to 10% of his brain cells were generally detected widely outside the brain, in tissues derived from endoderm, ectoderm, and mesoderm (Fig. 3C, fig. S13, and table S10), which suggests that these SNVs arose before the three germ layers segregated at gastrulation, yet still marked a minority of neurons.

Our data demonstrate that individual clones intermingle widely in the human cortex and that neurons from a given cortical region constitute

Fig. 1. Somatic SNVs are detected by single-neuron whole-genome sequencing. (A) Schematic of approach. Nuclei are isolated from a 0.5-cm³ piece of frozen post-mortem tissue by fluorescence-activated nuclear sorting (FANS), and DNA is amplified by Φ29 polymerase–mediated multiple-displacement amplification (MDA) and subjected to whole-genome sequencing (WGS). (B) Sample sequencing alignment tracks from brain B. One SNV is called uniquely in one sample (left), and one is shared between neurons 2 and 77 but absent from other single cells and heart (right). (C) Circos plot of SNV rate per megabase in brain B across human autosomes demonstrates that somatic mutations are distributed broadly. (D) SNVs per neuron are tightly correlated within each individual. (E) Most single-neuron SNVs in brain B are C>T transitions. Error bars in (D) and (E) denote SD; data points from individual neurons are spread horizontally for visibility.
Fig. 2. Somatic SNVs occur at loci that are expressed in the brain and associated with nervous system function and disease. (A) Coding exons are enriched for somatic SNVs. *P < 0.05, combined binomial test, Bonferroni-corrected. (B) SNVs in transcribed regions display a strand bias, suggesting that transcriptional damage influences the somatic SNV rate. *P < 0.05, analysis of variance with Sidak’s correction for multiple testing. Error bars in (A) and (B) denote SD. (C) Single-neuron SNVs correlate with epigenetic marks of transcription in the fetal brain and are depleted in heterochromatin, in opposition to the pattern observed in glioblastoma multiforme (GBM) SNVs. (D) Single-neuron SNVs occur in genes expressed in the cerebral cortex. Genes in the lowest expression quartile were significantly depleted, and those in the third quartile were significantly enriched, for single-neuron SNVs. *P < 0.05, +P < 0.05, Fisher combined one-tailed Poisson value for enrichment and depletion of SNVs, respectively. (E) Gene ontology categories associated with nervous system development and function are enriched for mutated genes across single neurons from brain B. (F) Examples of genes implicated in human disease that were mutated in single brain B neurons (for full list, see table S8). Green SNVs occurred in introns or downstream of the coding region, the orange SNV occurred in an exon and induced a missense mutation, and the magenta SNV is a nonsense mutation.

Table 1. SNVs found in 36 single human cortical neurons isolated from three normal individuals.

<table>
<thead>
<tr>
<th>SNV</th>
<th>Brain A (n = 10)</th>
<th>Brain B (n = 16)</th>
<th>Brain C (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Percentage</td>
<td>Count</td>
</tr>
<tr>
<td>Coding</td>
<td>23.7</td>
<td>1.3%</td>
<td>23.4</td>
</tr>
<tr>
<td>Silent</td>
<td>8.3</td>
<td>0.5%</td>
<td>7.1</td>
</tr>
<tr>
<td>Missense</td>
<td>14.1</td>
<td>0.8%</td>
<td>14.4</td>
</tr>
<tr>
<td>Truncating</td>
<td>1.3</td>
<td>0.1%</td>
<td>1.8</td>
</tr>
<tr>
<td>Noncoding</td>
<td>254.3</td>
<td>14.2%</td>
<td>243.5</td>
</tr>
<tr>
<td>Untranslated region</td>
<td>22.0</td>
<td>1.2%</td>
<td>21.0</td>
</tr>
<tr>
<td>Noncoding RNA</td>
<td>232.3</td>
<td>13.0%</td>
<td>222.5</td>
</tr>
<tr>
<td>Intronic</td>
<td>710.0</td>
<td>39.5%</td>
<td>660.1</td>
</tr>
<tr>
<td>Splice</td>
<td>3.0</td>
<td>0.0%</td>
<td>0.4</td>
</tr>
<tr>
<td>Other intronic</td>
<td>709.7</td>
<td>39.5%</td>
<td>659.7</td>
</tr>
<tr>
<td>Intergenic</td>
<td>805.0</td>
<td>44.9%</td>
<td>756.7</td>
</tr>
<tr>
<td>Mean</td>
<td>1793.0</td>
<td></td>
<td>1685.0</td>
</tr>
</tbody>
</table>

no fewer than five distinct clades of cells that trace their lineage back to separate mutation-marked pluripotent founder cells in the pre-gastrulation embryo. A given cortical neuron marked by SNV CI, for example, shares a more recent common cellular ancestor with a cardiomyocyte marked by this same somatic mutation than it does with approximately 75% of neighboring cortical neurons, with which no clonal connection is evident as far back as gastrulation. To confirm this polyclonal derivation, we dissected three consecutive 300-μm coronal sections from Brodmann area (BA) 40 and axially divided these sections into three regions, each 8 mm wide (Fig. 3D). Each region contained contributions from at least three, and usually four, of the major clades identified by single-neuron sequencing. Our data suggest that although late divisions of cortical progenitors likely generate neurons within a relatively restricted cortical zone (24, 25), a given anatomical column of cortex (26) derives from overlapping, intermingled clones from at least four, and likely more, developmental lineages (fig. S14).

Our results show that each human cortical neuron has a profoundly distinctive genome, harboring as many as 1458 to 1580 somatic SNVs, in addition to large CNVs and occasional retroelement insertions, as previously reported (5–8). These estimates are likely to improve with better sequencing and amplification methods as well as more samples. Similar SNV rates were seen at ages 15, 17, and 42, but whether older age is associated with increased SNV rates remains to be explored. These SNVs display signatures of mutagenic processes, such as transcription-associated DNA damage and a preponderance of meC>T
deamination. SNVs in coding regions of genes involved in nervous system development and mature neuronal function suggest a "use it and lose it" scenario, in which the very genes used for the function of a neuron are those most likely to be damaged during its life.

Our work demonstrates that somatic mutations can be used to reconstruct the developmental lineage of neurons, suggesting a potential "population genetics" of brain cells and representing a durable record of the series of cell divisions that gives rise to the human brain. This clonal mosaicism in the brain, an organ with exquisite arealization of function, may buffer the brain against deleterious clonal mutations that inevitably arise during development (2). Somatic mutations are also likely to modify the penetrance of germline neurological mutations to generate variable phenotypic effects of germline mutations in different family members, or even between identical twins.

REFERENCES AND NOTES
7. X. Cai et al., Cell Reports 8, 1280–1289 (2014).
INJURY RECOVERY

Function of the nucleus accumbens in motor control during recovery after spinal cord injury

Masahiro Sawada, Kenji Kato, Takeharu Kunieda, Nobuhiro Mikuni, Susumu Miyamoto, Hirotaka Onoe, Tadashi Isa, Yukio Nishimura

Motivation facilitates recovery after neuronal damage, but its mechanism is elusive. It is generally thought that the nucleus accumbens (NAc) regulates motivation-driven effort but is not involved in the direct control of movement. Using causality analysis, we identified the flow of activity from the NAc to the sensorimotor cortex (SMC) during the recovery of dexterous finger movements after spinal cord injury at the cervical level in macaque monkeys. Furthermore, reversible pharmacological inactivation of the NAc during the early recovery period diminished high-frequency oscillatory activity in the SMC, which was accompanied by a transient deficit of amelioration in finger dexterity obtained by rehabilitation. These results demonstrate that during recovery after spinal damage, the NAc up-regulates the high-frequency activity of the SMC and is directly involved in the control of finger movements.

M

otivation enhances and depression impedes functional recovery after neuronal damages (1, 2). However, the mechanism underlying this psychological effect on recovery remains obscure. The nucleus accumbens (NAc) plays a critical role in processing motivation to obtain reward (3–6). A large fraction of neurons in the NAc reflects the availability of a reward and the vigor of a subsequent approach to a target (7). However, the NAc is not thought to be directly involved in the control of motor functions such as finger dexterity. It has been shown that the NAc controls locomotion (8, 9); however, it might be through the regulation of vigor (10–12). On the other hand, in functional brain imaging studies in nonhuman primates, the activity of the NAc during a reach-and-grasp task was increased in association with the primary motor cortex (M1) during recovery from a partial spinal cord injury (SCI), which had not been observed before the injury (13, 14). These studies suggest that the interaction between the NAc and M1 plays a pivotal role in the recovery of finger dexterity. However, the causal relationship between the neural activities of M1 and the NAc and functional recovery remains obscure. In this study, we investigated the interaction between the neural activities of the NAc and M1.

After sufficient training on a reach-and-grasp task (Fig. 1A), four macaque monkeys were subjected to SCI, which was limited to the C4/5 level of the lateral corticospinal tract (Fig. 1B and fig. S3). Finger dexterity was immediately impaired in the three monkeys who had relatively large lesions (monkeys M, T, and R; Fig. 1B and fig. S3). Consistent with recent lesion studies (15–17), recovery of the precision grip exceeded 90% within 100 days in monkeys whose finger dexterity was impaired immediately after SCI. “Early-stage” recovery was defined as the period from SCI to the time point at which the success rate reached 100% for 2 successive days. “Late-stage” recovery was defined as the period after which the grip success rate remained repeatedly at 100% (Figs. 1B and 3B and movie S1). Finger dexterity was not markedly impaired in monkey D, presumably because the SCI was comparatively small (fig. S3).

To evaluate the potential of the NAc to motivate recovery, we determined the relationship between NAc and M1 activity during the two stages of behavioral recovery and compared with the pre-SCI stage. We examined oscillatory activity of the two regions while the monkey performed a reach-and-grasp task requiring dexterous finger movements at various time points before and after SCI in two monkeys (monkeys T and R, in which we conducted simultaneous recordings of electrocorticography (ECoG) in the sensorimotor cortex (SMC) and local field potentials (LFPs) in the NAc (Fig. 1C)). During the early-stage and late-stage recovery, time-frequency analysis of M1 activity showed long-lasting high-frequency gamma activity (115 to 400 Hz) through the period of reach-and-grasp movements, as compared with that observed before SCI (Fig. 1D, upper row and E). The high-frequency component of the NAc at the early stage appeared to have reached the similar level as the pre-SCI level (Fig. 1D, lower row and F). On the other hand, the low-frequency component in the NAc during the early recovery stage showed significant decrease during reaching and grasping (Fig. 1F). We successively employed Granger causality analysis (GCA) (18–20) to evaluate the directionality of interactions between M1 and NAc activity in the two monkeys. GCA demonstrated an increased signal flow from the NAc to M1 in the high-gamma band (115 to 400 Hz) during early-stage recovery (Fig. 2A, upper middle). In the late-stage recovery, the signal flow from the NAc to M1 was diminished (Fig. 2A, upper right). Compared with observations before SCI or during late-stage recovery, the signal flow from the NAc to the global SMC was much stronger (Fig. 2C).

Despite these analyses, it remained unclear whether the NAc is a critical modulator of SMC activity. We therefore injected muscimol at various time points before and after SCI to induce a temporary unilateral blockade of the NAc in the contralesional hemisphere. We tested the effects of NAc inactivation on finger dexterity in all four monkeys and on SMC brain activity in three