Genetic mosaicism in the human brain: from lineage tracing to neuropsychiatric disorders

Sara Bizzotto[1,2,3,4,5,6,7] ✉ *and Christopher A. Walsh*[1,2,3,4,5,6] ✉

Abstract | Genetic mosaicism is the result of the accumulation of somatic mutations in the human genome starting from the first postzygotic cell generation and continuing throughout the whole life of an individual. The rapid development of next-generation and single-cell sequencing technologies is now allowing the study of genetic mosaicism in normal tissues, revealing unprecedented insights into their clonal architecture and physiology. The somatic variant repertoire of an adult human neuron is the result of somatic mutations that accumulate in the brain by different mechanisms and at different rates during development and ageing. Non-pathogenic developmental mutations function as natural barcodes that once identified in deep bulk or single-cell sequencing can be used to retrospectively reconstruct human lineages. This approach has revealed novel insights into the clonal structure of the human brain, which is a mosaic of clones traceable to the early embryo that contribute differentially to the brain and distinct areas of the cortex. Some of the mutations happening during development, however, have a pathogenic effect and can contribute to some epileptic malformations of cortical development and autism spectrum disorder. In this Review, we discuss recent findings in the context of genetic mosaicism and their implications for brain development and disease.

Bulk variant allele frequency
Fraction of reads showing the mutant allele (compared with the reference human genome) calculated on the basis of the total number of reads covering the mutation position as obtained from deep bulk (non-single-cell) DNA sequencing.

Allelic imbalance
Where two alleles of a given gene are represented at different levels in the DNA sample, which results from higher amplification of one allele over the other at the time of whole-genome amplification and/or library preparation and consequently across sequencing reads.

✉*e-mail:* *sara.bizzotto@icm-institute.org*; *christopher.walsh@childrens.harvard.edu*

https://doi.org/10.1038/s41583-022-00572-x

Somatic DNA mutations (or variants) accumulate throughout development and ageing, resulting in no two cells in the human body sharing exactly the same DNA sequence, a phenomenon known as genetic mosaicism (GM). The advent of next-generation sequencing has brought the ability to explore the human genome on a large scale to identify genomic variation across many individuals, tissues and single cells. Some DNA variants are common in the general population[1]; others are rare and more likely associated with disease. Unlike germline mutations, non-inherited de novo mutations, such as somatic postzygotic mutations, are found in a subpopulation of related cells (clone) in the body (FIG. 1). The size of the clone depends on when the mutation occurs, with mutations occurring in early progenitors being inherited by many progeny, whereas late-occurring mutations are found in fewer cells. Somatic mutations can also be non-clonal when they occur in a single non-dividing, fully differentiated cell such as a neuron. Finally, somatic mutations in mitochondrial DNA also contribute to GM[2].

The study of GM in normal tissues has required significant technological advances that have had a huge impact in the field. Bulk DNA samples extracted from human tissue biopsy samples represent pools of thousands or millions of cells, each one contributing two copies of the genome. Thus, while a typical de novo heterozygous germline mutation has a bulk variant allele frequency (VAF) of ~50%, the VAFs of postzygotic somatic mutations are less than 50%, and for autosomal mutations (or mutations in the X chromosome in female individuals), VAFs represent half the number of cells carrying the mutation, referred to as mosaic fraction (MF) (FIG. 1). Identification or calling of somatic variants in bulk DNA sequencing is challenging because (1) high (more than 40%) VAF somatic mutations are easily confounded with germline mutations and (2) mutations with very low (typically less than 1%) VAFs require sequencing at very high depths and are difficult to discriminate from sequencing noise. Thus, identifying the lowest VAF mutations, and certainly those restricted to single cells, requires single-cell whole-genome amplification, which comes with issues related to culture condition artifacts, non-uniform genome amplification, allelic imbalance and the generation of technical artifacts due to polymerase synthesis errors and/or amplification bias. In recent years, some of the challenges associated with somatic variant calling have been overcome by (1) increasing homogeneity and accuracy in whole-genome amplification[3–5] and (2) generating sophisticated new

Variant calling
The process of identifying in sequencing data the sequence changes such as single-nucleotide variants and indels that are present in a given genome compared with the reference genome.

Indels
Genetic mutations described as insertions or deletions of one or more base pairs (typically fewer than 1,000) at a defined position compared with the reference human genome.

Copy number variants
(CNVs). Genetic mutations defined as deletions or repetitions of sections of the human genome of variable size that lead to partial aneuploidy (one copy instead of two copies) or multiploidy (more than two copies), respectively.

Inversions
Genetic mutations that consist of a section of the genome with the wrong (opposite) orientation, generated by a double break followed by change in the orientation of the DNA section and reinsertion in the same position.

Translocations
Changes in the location of sections of the genome that occur when parts of one chromosome are transferred to another chromosome.

Whole-chromosome gains or losses
Copy number variants of a type by which an entire chromosome is present in one copy instead of two copies (chromosome loss, aneuploidy) or alternatively when an entire chromosome is found in more than two copies (chromosome gain, multiploidy).

sequencing methods[6] and variant calling algorithms for deep bulk and single-cell sequencing data[7–13].

Single-nucleotide variants (SNVs) are the most frequent variants in the human genome, followed by small (1–4-bp) insertions or deletions (also referred to as indels). Genomic structural rearrangements or structural variants, however, such as large (more than 1 Mb) copy number variants (CNVs), inversions, translocations, whole-chromosome gains or losses and mobile genetic element insertions, are also found at the somatic state[14–17]. Although less frequent, they have a higher probability of causing severe disease, as they typically involve many genes. Detection of structural variants, and especially those with low VAF, is computationally more challenging than detection of SNVs and small indels. Even though long-read sequencing is now offering more suitable data[18], and better computational algorithms are being developed that call somatic CNVs and transposon insertions[15,19–22], most of the existing structural variant calling and annotation methods are suitable only for high MFs found in cancer, an issue that needs to be addressed.

Although somatic mutations have long been studied in the context of cancer, these technologies have recently allowed the exploration of somatic mutation rates and patterns also in other types of diseases and in normal tissues[23–39], revealing novel insights into human development, physiology and pathology. Clonal somatic mutations, and especially SNVs that occur with high frequency during development, have been used as a tool to perform retrospective lineage tracing in humans[38–44]. Here, we discuss current knowledge of the rates and mechanisms of accumulation of somatic variants in the brain. We explain how GM has created a whole new framework for the study of brain development directly in humans, and cover what GM has revealed so far about the clonal architecture of the human brain. We intersect normal development with the role of somatic mutations in developmental neuropsychiatric diseases. Finally, we discuss the major technical limitations in the field, and what we think are some of the most promising future directions to follow.

Mechanisms of brain genetic mosaicism

Somatic mutations can occur from the very first postzygotic cell division onwards[39,44–46], and mutations that occur during development contribute substantially to the somatic variant repertoire of the human brain[39] (FIG. 2). Whole-genome sequencing (WGS) of several human tissues estimated that ~2.4–2.8 variants accumulate at each of the first two or three postzygotic cell divisions[38,44]. Other studies in pairs of monozygotic twins and in adult brains suggested 3–3.4 mutations distinguish the first two postzygotic lineages[45,46]. Thus, overall, a typical individual would accumulate a total of ~100 genome-wide somatic SNVs during the first five cell divisions after conception, whereas the mutation rates are thought to decrease to ~1–2 mutations per cell division from the eight-cell stage onwards[38,40,47]. Fixed double-stranded somatic mutations often result from incorrectly repaired DNA damage caused by cellular respiration and exposure to environmental stressors[48]. However, mutations also occur during DNA replication, transcription and epigenetic modifications, all processes that are highly active during development. The short cell cycle and the reduced expression of G1 checkpoint proteins that verify DNA integrity[49] might explain the higher mutation rates at the earliest stages.

Once the various tissues start to differentiate, the total numbers and accumulation rates of somatic variants reflect the unique replication and/or transcription behaviours of the cells within those tissues. Although there are no specific data regarding somatic mutation rates during gastrulation and in neuroectoderm, the precursor of the CNS, sequencing of clones expanded from fetal neural progenitors at 20 weeks of gestation, a period of active cortical neurogenesis that follows gastrulation and establishment of the major structures of the CNS, found that each neural progenitor carries 200–300 somatic SNVs, with an accumulation rate of ~5.1 SNVs per day (~8.6 per cell division)[50]. Thus, this study suggests that during neurogenesis, neural progenitors in the developing brain seem to acquire somatic mutations at higher rates than earlier embryonic progenitors.

Postmitotic neurons in the adult brain continue to accumulate somatic mutations throughout life. Single neuronal genomes obtained from individuals spanning 5 months to 82 years old and amplified by multiple-displacement amplification showed 300–900 somatic SNVs per neuronal genome at ~1 year of age, which increased linearly at a rate of 23–40 per year to reach ~1,500 in adult individuals, and ~2,500–4,000 somatic SNVs per genome in individuals in their 80s[34,41]. Neuronal genomes amplified by a different method, called multiplexed end-tagging amplification of complementary strands, had an accumulation rate of ~16 SNVs per genome per year[5], similar to the ~16 somatic SNVs per genome per year calculated using primary template-directed amplification[13]. Another recently introduced sequencing method, called NanoSeq, found a linear accumulation rate of 19.9 SNVs per neuronal genome per year, with SNVs increasing from 800–1,000 in adults to 1,600–2,000 in individuals older than 80 years[6]. Thus, although the rates and numbers of somatic SNVs are being refined as different technologies for whole-genome amplification, whole-genome sequencing and variant calling are developed[5,6,13], these studies are all concordant in showing that neurons in the ageing human brain accumulate somatic SNVs at rates of ~15–20 per genome per year.

Although information is less abundant, indels have also been shown to contribute to the somatic

Author addresses

[1]Division of Genetics and Genomics, Manton Center for Orphan Disease Research, Allen Discovery Center for Human Brain Evolution, Boston Children's Hospital, Boston, MA, USA.
[2]Department of Paediatrics, Boston Children's Hospital, Boston, MA, USA.
[3]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA.
[4]Department of Paediatrics, Harvard Medical School, Boston, MA, USA.
[5]Department of Neurology, Harvard Medical School, Boston, MA, USA.
[6]Broad Institute of MIT and Harvard, Cambridge, MA, USA.
[7] Sorbonne Université, Institut du Cerveau (Paris Brain Institute) ICM, Inserm, CNRS, Hôpital de la Pitié Salpêtrière, Paris, France.

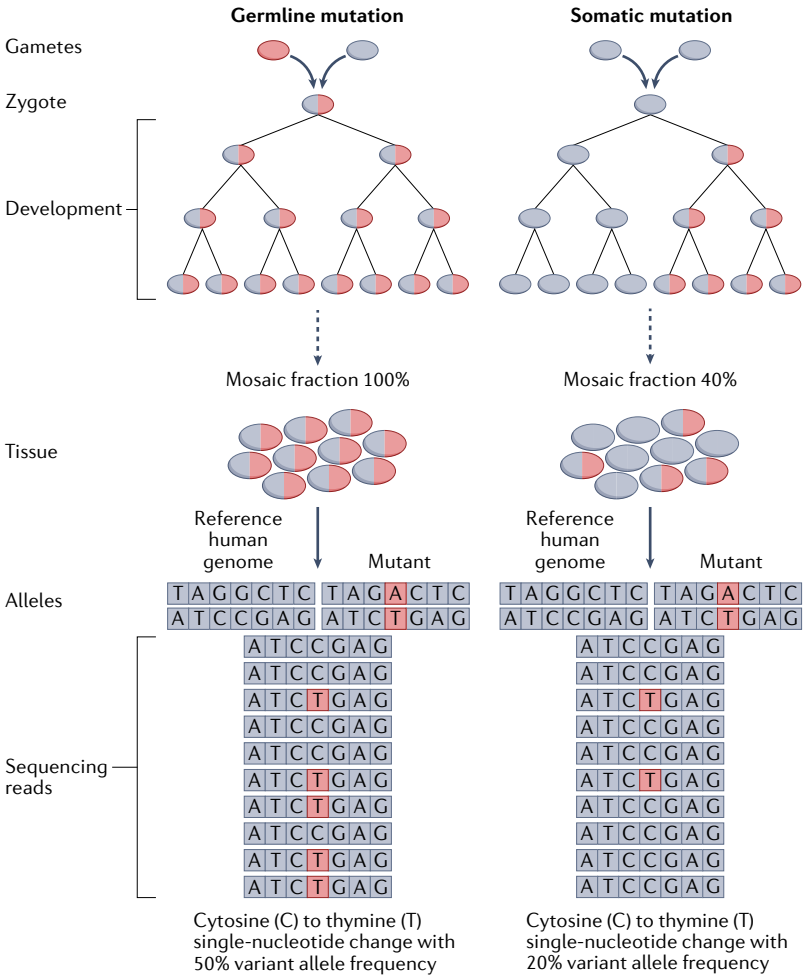

Fig. 1 | **Detection of somatic mutations in deep sequencing of bulk DNA.** Germline de novo mutations (left) occur in one of the two haploid gametes (orange) that form the diploid heterozygous zygote, and thus pass on to the descending progeny. Somatic mutations (right) occur during development in a progenitor cell that will pass the mutation to all the descending progeny, which will be heterozygous for the mutation. Somatic mutations can also occur in postmitotic cells (not shown) and can be restricted to a single cell in the body. Bulk DNA samples represent pools of many single cells. In the case of a heterozygous germline mutation, 100% of the cells contribute one mutant allele (compared with the reference human genome), which is found in ~50% of the sequencing reads. Postzygotic somatic mutations are carried by a fraction of the cells in the tissue (for example, 40%). Thus, the mutation is found in a number of reads that corresponds to half the mosaic fraction, 20% in the example. It is important to note that this is the case for the great majority of human somatic mutations, except for those occurring in the sex chromosomes of a male human. In this case, the mosaic fraction corresponds to the percentage of mutant reads (variant allele frequency).

Mobile genetic element insertions
Insertions in the genome of new copies of discrete segments of genomic DNA.

Long-read sequencing
Also known as third-generation sequencing, a class of DNA sequencing methods that generate reads much longer (10–15 kb, up to 30 kb) than classical short-read next-generation sequencing methods (~150 bp).

variant repertoire of the ageing human brain[6,13]. NanoSeq reported 150–200 indels in young adults and 250–300 in older individuals, for an accumulation rate of 2.5 indels per neuron per year, similar to the rates of somatic indels revealed by primary template-directed amplification[6,13]. Overall, although somatic SNVs seem to accumulate in neurons at rates comparable to those in some other differentiated cells[27,30,31,37,43], neurons seem to contain a higher proportion of indels than other tissues[6]. Although much less frequent than SNVs, somatic CNVs and retrotransposition events were also characterized in neurons: 10% of neurons are estimated to carry at least one megabase-scale de novo somatic CNV[51–53], whereas 0.2–1 long interspersed nuclear element (L1) retrotransposon insertions per genome were observed[16,54,55]. A more recent study estimated that 7.9% of healthy adult brains may carry at least one clonal somatic CNV[56].

In addition to mutation rates, the genomic distribution, the combination of substitution types and their trinucleotide context also point to mechanisms of somatic SNV accumulation. In cancer tissues, somatic mutations usually correlate with DNA replication timing, indicating that DNA replication plays a major role in mutagenesis[57]. Similarly, cell replication during development may contribute to somatic mutations by DNA synthesis errors and/or by errors introduced during the fixing of unrepaired single-strand DNA damage (FIG. 2a). Signatures are characteristic combinations of somatic mutation types that are associated with specific mutational processes. The Catalogue of Somatic Mutations in Cancer (COSMIC)[58] is a database of somatic mutation signatures found in cancer, and thus associated with specific known causes. Signatures newly identified in a given tissue can be compared with the COSMIC reference signature database to understand the contribution of different known mechanisms of somatic variant generation to normal tissues. Somatic SNVs in the early embryo are mostly C>T transitions and are particularly enriched in CpG motifs, sites of DNA methylation[39,40,46]. These mutations are mostly due to unrepaired G:T mismatches in double-stranded DNA generated by the endogenous process of spontaneous or enzymatic deamination of 5-methylcytosine to thymine[39,40,44,47,50,59] (FIG. 2a). This is the proposed cause associated with COSMIC signature 1 (SBS1), a cell division and mitotic clock signature that correlates with age in cancer. The same mechanism was also observed in clones derived from human fetal neural progenitors, with early high-mosaic mutations contributing more C→T transitions to the total mutation set[50] than later low-mosaic mutations. The latter, on the other hand, contributed more C>A transversions, pointing to oxidative damage as an additional contributing cause in neural progenitors[50] (FIG. 2b).

Cell replication in the adult brain is associated mostly with glial cell turnover and does not play a role in the accumulation of somatic mutations in postmitotic fully differentiated neurons. Although a thorough assessment of somatic mutations in adult glial cells is yet to be done, a recent study found a greater prevalence of oncogenic clonal somatic mutations in the white matter than in the grey matter of normal adult brains. The enrichment of these mutations in the glial cell population suggests that at least some of the somatic events that drive tumour formation may occur during postnatal clonal expansion of glial cells[56]. By contrast, in adult prefrontal cortex neurons, the non-clonal somatic SNV genomic distribution did not correlate with replication timing, but instead somatic SNVs were enriched in coding exons with a strand bias indicative of transcription-associated mutation[6,13,41] (FIG. 2c). In contrast to neural progenitors, where a significant depletion of somatic variants was observed in open chromatin regions[50], two studies found that somatic SNVs in mature neurons correlated with chromatin markers of transcription and were enriched in highly expressed genes[13,41]. NanoSeq and

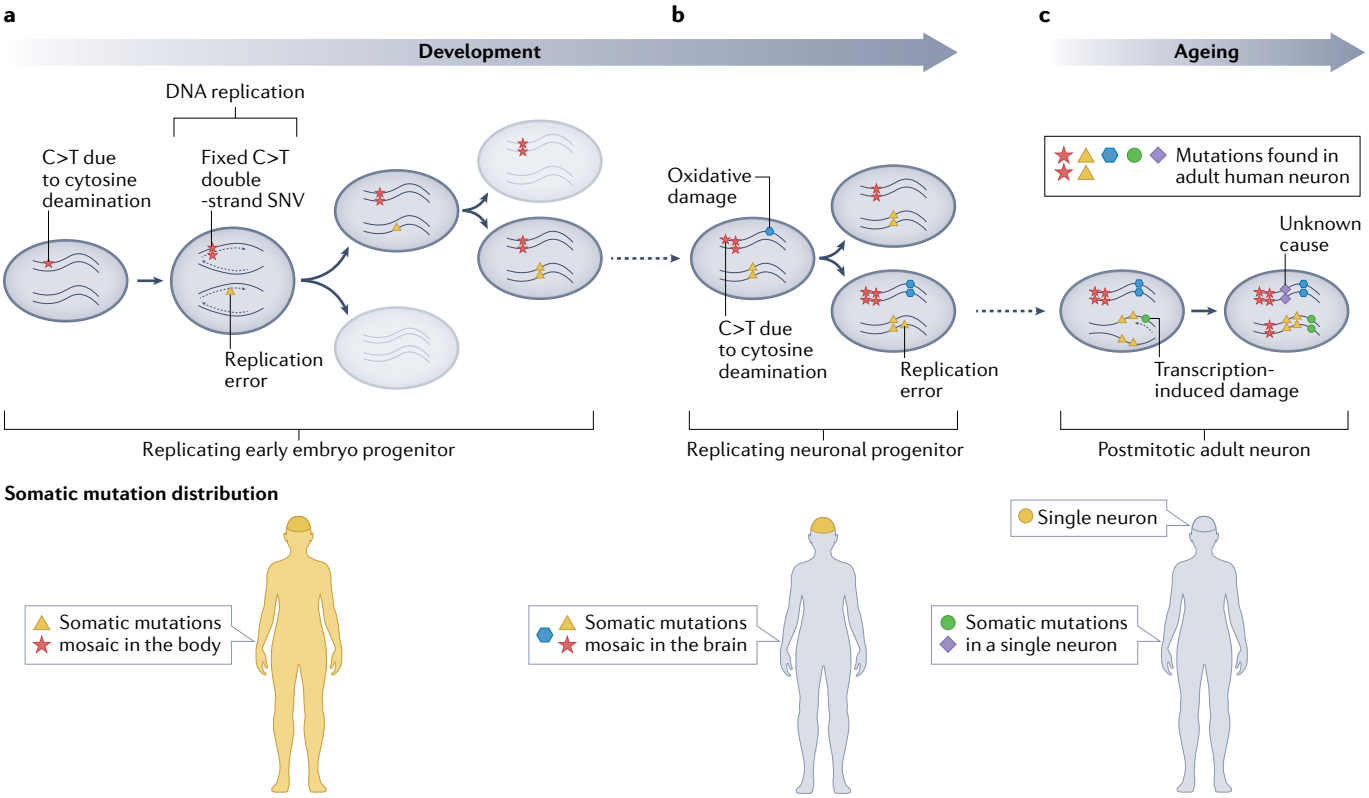

Fig. 2 | **Mechanisms of genetic mosaicism in the human brain. a** | Somatic variants that accumulate in early postzygotic embryonic progenitors are associated mainly with cytosine deamination and with DNA replication, and are found as a clonal mosaic across the whole body (yellow shadow in the lower panel). Events of spontaneous deamination of cytosine to thymine that are not repaired before DNA replication become fixed C→T double-strand single-nucleotide variants. Furthermore, errors during DNA replication that are not correctly repaired can propagate with subsequent cell divisions and become somatic mutations. **b** | Although replication and cytosine deamination are still a significant contributor to somatic mutations in neural progenitors, oxidative damage has also been proposed as a significantly active mechanism in these cells. In this case, fetal neural progenitor mutations would be found as a clonal mosaic only in the brain, carried by all the neural cells derived from the neural progenitor where the mutation first occurred. **c** | An adult human neuron carries all the clonal variants accumulated during development (panels **a** and **b**), and an additional set of non-clonal variants that are mainly associated with transcription and other mechanisms of unclear cause.

primary template-directed amplification also detected an enrichment of indels in highly expressed neuronal genes[6,13]. Consistent with these observations, two recent studies mapped DNA repair activity across the genome and found that brain-associated enhancers and genes with neuronal function are hotspots for DNA repair, indicative of increased damage in these regions[60,61]. In neurons, a 'clock-like' signature that increases with age was identified made mostly of C>T and T>C transitions. This signature resembles COSMIC signature 5 (SBS5)[6,34], which appears ubiquitous in both normal and cancer tissues[37] but has an unclear underlying cause. Thus, although transcription seems to contribute to somatic variants in postmitotic neurons, other mechanisms are likely to play an important role but remain poorly understood (FIG. 2c).

Genetic mosaicism and lineage tracing

Lineage tracing is the reconstruction of the nested cell divisions and clonal relationships between cells of an organism. For cell divisions to be effectively followed, related cells need to carry a unique lineage identifier such as a short nucleotide sequence (or barcode). Recently, techniques have been developed that take advantage of CRISPR–Cas9 genome editing or transposon tagging to perform in vivo barcoding by artificially introducing stochastic genetic variants (or scars) in cycling cells[62–66]. High-throughput next-generation sequencing is then used for the massive parallel sequencing of hundreds or thousands of cells and the building of prospective lineages. These elegant approaches have tracked the development of entire model organisms, such as zebrafish[63,65,67], *Caenorhabditis elegans*[68] and the mouse[64,69,70], but a comprehensive study in the mammalian brain has yet to be performed[71], and prospective lineage tracing in humans is limited by the availability of viable tissues where these technologies can be applied[72]. GM, however, is a natural barcoding system present in humans that has the potential to reveal the clonal structure of the tissues by allowing retrospective reconstruction of cell divisions[39].

Early human embryo. GM applied to lineage tracing has already revealed surprising aspects of early steps of human development starting with the first postzygotic cell division. Somatic variants detected in human blood and other adult tissues, placenta and DNA from monozygotic twins have shown that human development is an asymmetric process[35,38–40,42,43,45–47]. The VAF distribution

Retrotransposition events
Mobile genetic element insertions of a type involving class I transposable elements that are able to copy and paste themselves by generating an RNA intermediate.

Gyrencephalic
Condition by which the surface of the cerebral cortex is characterized by the presence of convolutions made of alternating gyri and sulci, in contrast to the lissencephalic cerebral cortex, where the surface is smooth.

of somatic mutations identified in blood first suggested that the first two daughter cells of the human embryo contribute with a 2:1 asymmetry to the blood lineage[44]. A higher imbalance of 90:10 was found in blood, and imbalances of 70:30 and 80:20 were found in cells from the urinary tract from two individuals[42]. By sequencing brain tissue in more than 70 individuals, we showed that the contributions of the first two postzygotic progenitors to the brain are variable and range from 50:50 symmetry to 80:20 asymmetry or potentially even higher[39]. A more recent study confirmed the existence of variabilities across tissues and individuals, and found a range of asymmetries between 60:40 and 93:7 (REF.[38]). These asymmetries are likely the consequence of bottlenecks of stochastic clonal selection during early processes such as blastula formation, when embryonic and extra-embryonic tissue lineages separate[35,39,40,42,43,45–47].

Mutation MFs can estimate cell population size at the time of variant occurrence as approximately 1/MF. Following this rule, three of the eight-cell stage progenitors were suggested to give rise to the inner cell mass[38]. Similarly, somatic mutation MFs indicate blastocyst formation at the 4-cell to 16-cell stage in humans[40], while twinning is thought to happen after five or six postzygotic cell divisions (32–64 cells)[45]. The blood lineage was suggested to derive from ten inner cell mass founder cells[44], whereas our analyses of somatic SNVs across germ layers calculated ~170 epiblast cells later at the start of gastrulation[39]. The concordance of these estimates based on MFs with cell counts in in vitro studies of human embryonic tissue up to the pregastrulation stages[73] confirms their accuracy.

GM has been applied also to later developmental steps in human brain formation. The CNS is derived from neuroectoderm forming the neural tube, and lineage tracing in barcoded mice showed that commitment to the CNS anteroposterior axis is established before the left–right axis[69]. Consistent with this, early embryonic lineages marked by somatic variants until the fourth postzygotic cell generation showed significant correlation between the VAFs and the anteroposterior location of the brain region, suggesting that early progenitors may already be specified to allocate their daughter cells in the brain according to an anteroposterior axial rule[42]. Besides this, in our recent study of mosaic mutations across several tissues, including the CNS, we estimated that the human forebrain forms from ~50–100 founder neuroectoderm progenitors[39].

Human cortical development. The cerebral cortex is the largest forebrain structure and in humans is made up of ~100 billion neurons and glial cells born from gestational week 5 to gestational week 28 and spatially organized into an intricate six-layer structure. In mammals, progenitors located in the ventricular zones of the dorsal forebrain give rise to all the major types of excitatory neurons, which migrate radially to form the cortical layers in an inside-out order. Inhibitory interneurons are instead born in ventral forebrain proliferative niches called ganglionic eminences and migrate tangentially to reach the cortex[74]. Several elegant strategies involving retrovirus transfection[75,76] or mouse genetic engineering for recombination-mediated expression of marker genes[77–86] have investigated lineages and clonal patterns in the mouse brain[87], showing defined unitary output in neuronal production and organization of clones in vertical clusters[84]. However, very little is known about the clonal structure of the human brain, and more recent approaches suggest that excitatory neuron production from dorsal progenitors may be more complex even in the mouse, with progenitor cells generating lineages that exhibit a wide range of sizes and laminar configurations[71]. Retroviral labelling of neuronal progenitors in the ferret and primate neocortex suggests that related cells might disperse widely in more complex gyrencephalic brains[88–90], where additional proliferative layers are responsible for a much higher neuronal production[91,92]. Thus, although rodent models have provided key knowledge of cortical development, the complexity and expansion of the human brain implies the existence of human-specific mechanisms that need to be addressed.

The abundance and spatial topography of variants that mark early embryonic progenitors give key insights into cortical development. Linking mosaicisms in the cortex with reconstructed early embryo lineages, we recently showed that pregastrulation progenitors at the eight-cell stage contribute asymmetrically to the cortex, with the average contribution of the different clades ranging from 30% to less than 5% of the total cells[39] (FIG. 3). Whether this variability reflects an asymmetric repartition of the clones at the time of specification of founder cortical progenitors or differences in proliferation rates of neural progenitors remains to be determined. Analysis of multiple samples spanning different regions of the cortex showed that the VAF of the same SNV often varies considerably from one region to another[38,39,41], suggesting differences in the degree of local amplification of the same clone.

Mosaic variants that first occur in neural progenitors are inherited only by their daughter cells in the brain, and are thus specifically informative for neural progenitor behaviour. These variants are usually found in less than 2% of the cells sampled across the whole rostrocaudal cortex[16,39,41]. Topographic analyses of SNVs marking progressive early-to-late lineages showed that a gradual spatial restriction seems to occur where early lineages span the entire rostrocaudal cortex, whereas late sublineages appear restricted to specific regions[16,39]. Several studies have identified a few examples of mosaic variants restricted to the frontal lobe and not detectable in posterior areas of the cortex[16,39,41] (FIG. 4). This phenomenon may suggest that the frontal lobe — separated from the rest of the cortex by the sylvian fissure and the central sulcus — might have a slightly different clonal structure[39]. Although some clonal variants may have unexpectedly high MFs due to local clonal expansion of glial cells[56], clones showing a spatial restriction to one region of the cortex usually represent less than 1% of the cells in that specific region[16,39]. For example, a somatic retrotransposition event with very low mosaicism (0.09%) was found only in five adjacent regions restricted to the left middle frontal gyrus of

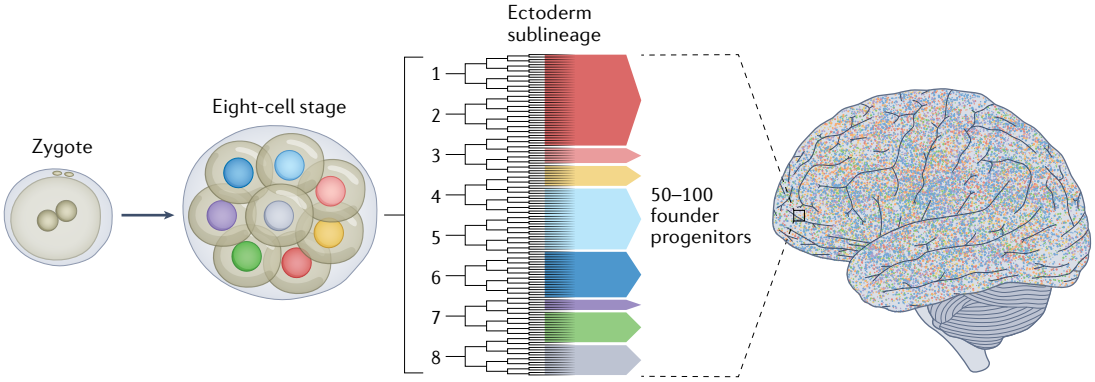

Fig. 3 | **The human brain is a clonal mosaic.** After fertilization, the zygote divides to reach the eight-cell stage. Sublineages derived from these first eight progenitors of the human embryo differentiate into ectoderm and, subsequently, contribute asymmetrically to the brain, with some lineages contributing more (for example, red clone) and others contributing much less (for example, yellow and pink clones)[39]. This unequal contribution may be due to bottlenecks of stochastic clonal selection during development and/or may be due to different proliferation rates between clones. The ectoderm later generates 50–100 founder progenitors of the forebrain[39]. The adult human cerebral cortex is a mosaic of clones traceable to the eight-cell stage, with a cortical patch containing intermingling neurons of different clonal derivation traceable to the eight-cell stage.

the cortex[16], and similar degrees of spatial restriction were observed for other clonal variants[39]. The presence of such low-mosaicism variants across at least 2 cm of the cortex suggests a quite wide dispersion of related cells, and that within any given radial patch of the cortex there must be many intermingling clones (FIG. 3), which implies additional complexity.

A step further in the understanding of the establishment of clonal patterns in the brain concerns how lineages contribute to different cell types. Indeed, excitatory neurons, interneurons and different types of glial cells are known to behave differently during development in terms of specification timing, proliferation and migratory patterns[93]. The simultaneous analysis of somatic DNA variants and markers of cell types, typically RNA expression, is required to discriminate lineages of distinct cell types but is not yet possible at high throughput. We recently used cell sorting to show that early embryonic lineages contribute unequally to the two broadest classes of cells in the cortex, neuronal (NeuN-positive) and non-neuronal (NeuN-negative) cells, represented by astrocytes, oligodendrocytes, microglia, oligodendrocyte precursors and endothelial cells[39]. Spatial analyses of cortical clones coupled with sorting of NeuN-positive neurons showed that although a clone restricted to the middle frontal gyrus was present in neuronal cells and absent from non-neuronal cells, another, more expanded clone was found in both cell populations[16]. This may indicate that whereas the former clone derived from a neural progenitor with no gliogenic potential, the latter was probably the product of an earlier progenitor that gave an output of both neurons and glia.

Parallel RNA and DNA analysis after deep sequencing (PRDD–seq) is a technique that combines targeted sequencing of clonal DNA variants with gene expression analysis of a subset of cell type-informative genes. This technology has allowed the tracking of lineage markers in the human cortex and the distinguishing of excitatory and inhibitory neuron subtypes[94]. Early lineage markers were found in both excitatory and inhibitory neurons, while later nested clades were progressively enriched in the excitatory-neuron population, confirming that the two main neuronal subclasses are generated mainly from distinct progenitor pools in humans as they are in non-humans. Additionally, the progressive restriction of later markers in cells classified as middle-layer and upper-layer excitatory neurons was coherent with the inside-out development of cortical layers from dorsal excitatory neuron progenitors. As excitatory neuron-restricted clades were present in less than 4% of the cells sampled from a cortical volume of less than 5 mm^3, and considering that 40% of neuronal cells in the cortex are excitatory neurons, this study also suggested at least ten excitatory neuronal progenitors contributed to such cortical volume[94], which should be traceable back to the eight-cell stage (FIG. 3).

Interneurons in the human brain are much more numerous and appear more diverse than in rodents, and very little is known about their developmental lineages. Although a recent study showed that human dorsal cortical progenitors are able to generate both excitatory and inhibitory neurons[95], interneurons derive mostly from ventral proliferative niches in the ganglionic eminences, with more than half of the inhibitory neurons in the human cortex likely originating from the caudal ganglionic eminence[96]. Studies have suggested that SST-positive and PVALB-positive interneurons, derived mainly from the medial ganglionic eminence, preferentially populate deep layers of the cortex, whereas interneuron subtypes generated in the caudal ganglionic eminence, which are LAMP5 and VIP positive, tend to occupy upper layers[97–99]. On the basis of PRDD–seq, PVALB-positive, LAMP5-postive and SST-positive interneurons do not show a detectable inside-out pattern of development in the human brain[94]. Thus, at this time there is no definitive information about patterns of interneuron clones and inside-out development[100,101]. This certainly is an interesting topic for future exploration that requires sampling larger numbers of cells for lineage tracing.

Cytomegalic dysmorphic neurons
Abnormal neurons that represent a histopathological hallmark of focal cortical dysplasia type 2 and are characterized by a significantly enlarged cell body and nucleus, misorientation, abnormally distributed intracellular Nissl substance and cytoplasmic accumulation of neurofilament proteins.

Developmental brain disorders

Although the great majority of somatic variants in normal individuals appear to be innocuous[39], some somatic mutations are deleterious, such as those that occur in sensitive genomic regions such as tumour driver genes[24,25,27,30–32,36,102] or genes and regulatory elements involved in brain disorders such as neurodegenerative disorders[103–107] and neuropsychiatric disorders[15,46,108,109]. In the human brain, the rates of early somatic SNVs in exons and areas of open chromatin are higher than expected from a uniform mutation rate, suggesting that coding regions are particularly vulnerable during development[46]. It has been estimated that each individual can accumulate ~1.9 exonic somatic SNVs during the first five cell divisions of development, of which ~45% are potentially damaging (for example, missense, stop codon, frameshift and splicing altering)[110]. Estimates based on analysis of fetal brains (gestation week 5 to gestation week 21) suggest that ~3% of somatic SNVs may have functional consequences, with each progenitor cell at gestation week 20 carrying about 12 non-benign somatic SNVs[50].

While the cumulative action of non-clonal somatic mutations is thought to play a role in neurodegeneration and ageing[6,34,107], pathogenic somatic mutations that occur during development have an established role in cortical malformation during development[111–113]. Somatic SNVs and CNVs are thought to play a role in schizophrenia[14,114,115], but most existing studies are limited by cohort size. More is known about somatic mutations in epileptic focal cortical dysplasia (FCD) and autism spectrum disorder (ASD)[116–139].

Focal cortical dysplasia. The role of somatic mutations is well established for FCD spectrum disorders[116], including tuberous sclerosis complex and hemimegalencephaly (HME). Somatic pathogenic gain-of-function mutations in genes coding for activators of the mechanistic target of rapamycin (mTOR) pathway, such as *AKT3*, *PIK3CA*, *RHEB* and *MTOR* itself, can cause FCD type 2 (FCD2)[116] and HME[117–122]. Double-hit loss-of-function mutations, one germline and one somatic, were found in repressor genes of the mTOR pathway, such as *TSC1*, *TSC2*, *DEPDC5*, *PTEN*, *NPRL2* and *NPRL3*, and can lead to FCD and HME, as well as tuberous sclerosis complex[121,123,124]. Overall, mTOR pathway mutations account for as many as 50–60% of FCD2 and HME cases[121]. Somatic mutations in the *SLC35A2* gene were also recently implicated in atypical FCDs defined as mild malformations of cortical development with oligodendroglial hyperplasia in epilepsy[125,126]. mTOR pathway hyperactivation increases cell metabolism and biosynthesis, leading to the presence of abnormal cytomegalic dysmorphic neurons

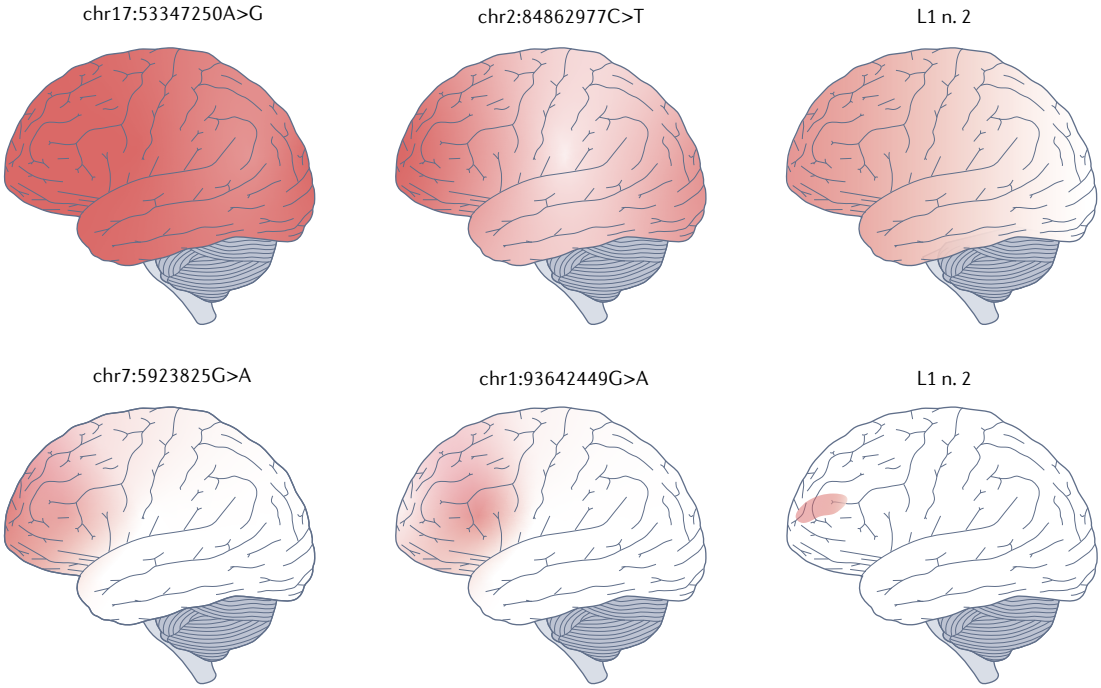

Fig. 4 | **Spatial distribution of somatic variants marking human brain clones.** Examples of clonal somatic variants derived from studies that characterized mosaic fractions (MFs) across the cerebral cortex by ultra-deep sequencing of bulk DNAs derived from multiple locations. In red are the areas of the brain where variants were detected, with gradients representing MFs. An early brain-restricted variant (chr17:53347250A>G) marking founder forebrain progenitors was found all over the cortex in ~2% of the cells on average[39]. Chr2:84862977C>T was also found all over the cortex, although at a slightly lower average MF of 1.3%. MFs were higher in the prefrontal cortex and occipital lobe (maximum MF 3.8%) and lower in medial regions around the central sulcus (minimum MF 0.03%)[39]. Long interspersed nuclear element-1 n. 2 (L1 n. 2) showed a gradient of enrichment, with higher MFs detected in the frontal lobe (1.7%) and decreasing to a minimum of 0.01% as we move towards the occipital and temporal lobes[16]. Chr7:5923825G>A and chr1:93642449G>A were both restricted to the frontal lobe[39], as was Long interspersed nuclear element-1 n. 1 (L1 n. 1), which was found only in an area of the middle frontal gyrus spanning 2 cm × 1 cm at MFs of 0.04–0.2%[16].

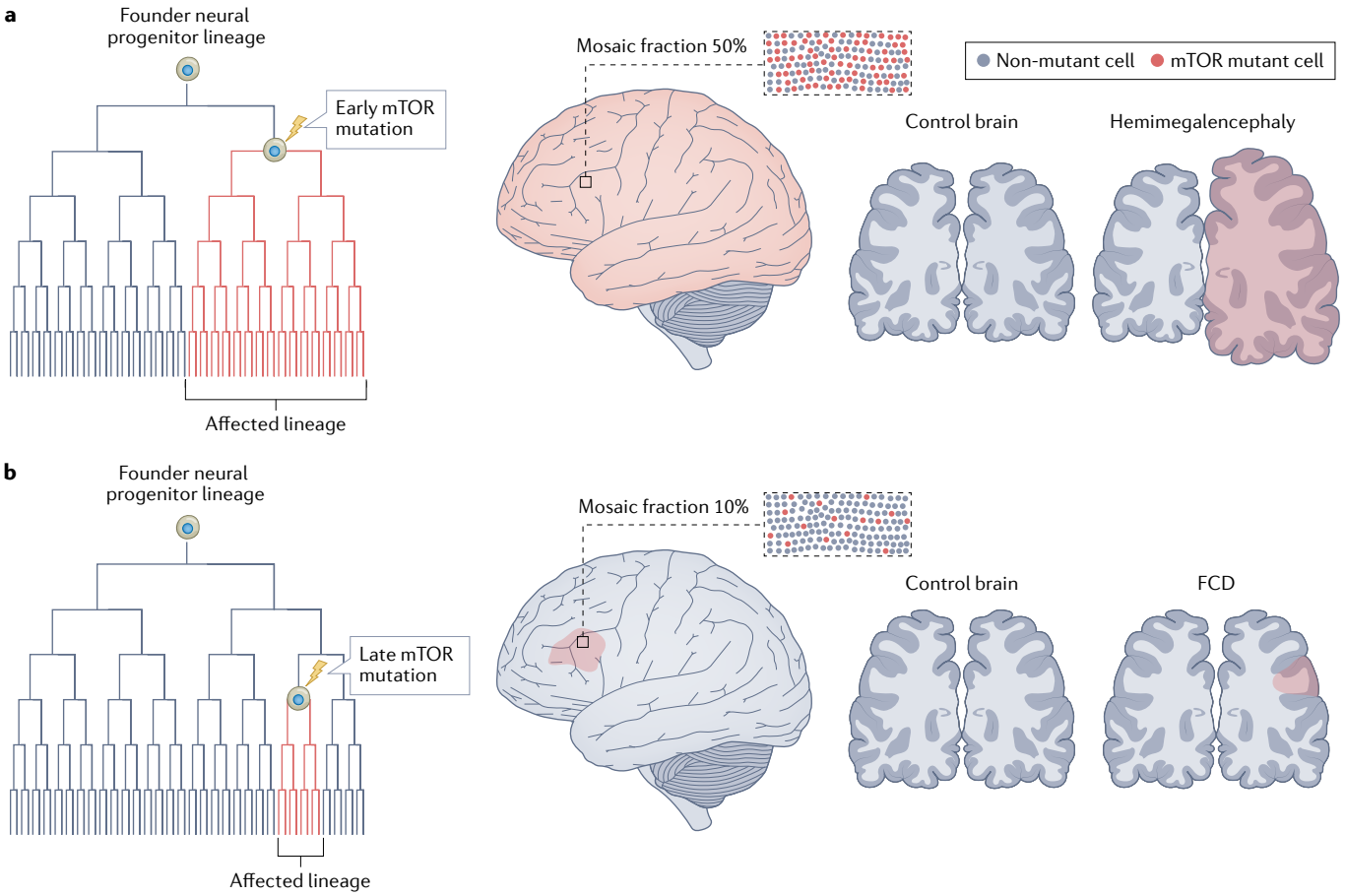

Fig. 5 | **Pathogenic somatic mutations in FCD spectrum disorders. a** | A somatic pathogenic mutation in mechanistic target of rapamycin (mTOR) pathway genes occurring in an early neural progenitor affects an entire cortical hemisphere, where it is found in 50% of the cells (variant allele frequency 25%), leading to hemimegalencephaly. **b** | A somatic mutation in mTOR pathway genes occurring in a late neural progenitor affects a smaller region of the cortex, where it is found in 10% of the cells (variant allele frequency 5%), leading to a localized focal cortical dysplasia (FCD).

and balloon cells in the cortical tissue, histopathological hallmarks of FCD2 and HME[116,127]. The density of these abnormal cells in FCD2 and HME lesions usually correlates with the size of the affected area and the mutation VAF[117,121]. Indeed, laser microdissection of morphologically normal and abnormal cells in FCD2 and HME and genotyping of enriched populations have indicated that abnormal cells preferentially carry the mutation[121], although not exclusively.

The developmental timing of mutation occurrence and the effect of the mutation on cell proliferation has an impact on the lesion size and the severity of the disease. FCD somatic mutations are usually undetectable in blood and thus seem in most cases restricted to the cortex. This implies that they occur later than the establishment of the founder progenitors of the forebrain[39]. Mutations leading to HME presumably occur in early cortical progenitors after the establishment of the left–right axis (FIG. 5a). Smaller FCDs, however, are likely due to mutations that hit a later neural progenitor (FIG. 5b). Although current detection limits make it difficult to call mosaic SNVs with VAF less than 1% and structural variants, mutations in FCD and HME lesions often have VAFs higher than expected for a spatially restricted clone[16,39,117], suggesting a positive selection and expansion of the mutant clone; current detection limits might partially explain the 40–50% of genetically unsolved cases. In HME, VAF ranges from 5% to 30%[117], which in normal lineages is usually seen in variants arising in the very first postzygotic divisions, even before blastula formation[39,42]. Thus, mTOR pathway mutations with VAF above the detection limits seem to confer a proliferative advantage, which increases the density of affected cells in the lesion. Mutations that are present in such a small number of cells to be undetectable without prior isolation of the mutant cells might occur later enough to have little effect on clonal expansion.

Studies of neuronal and non-neuronal cell populations sorted from human surgical resections, together with functional studies in animal models, have shown that neurons are certainly involved in FCD pathology[117,128,129]. Indeed, although both neuronal and non-neuronal cells carry the pathogenic mutation in the biggest FCD lesions and HME, in small FCD lesions with VAF less than 5%, by contrast, mutations seem to be carried mostly by neuronal cells[117]. Although we cannot rule out whether few glial cells do exist that carry the mutation in these lesions, the data suggest that the

Balloon cells
Abnormal cells of unclear identity that represent a histopathological hallmark of focal cortical dysplasia type 2B and that are characterized by a large cell body, opalescent glassy eosinophilic cytoplasm (visible by haematoxylin and eosin staining) and absence of Nissl substance.

Positive selection
The process by which a clone acquires a selective advantage and proliferates more with respect to surrounding clones, leading to an over-representation in the tissue.

mutation may have occurred in a progenitor that gave rise mostly if not exclusively to neurons[93]. In conditional mouse mutants for *PIK3CA*^{H1047R}, when the excitatory neuron lineage is affected (with use of *Emx1*–Cre as a driver), the phenotype in the cortex is severe, characterized by dramatic megalencephaly and cortical dyslamination[117]. By contrast, when the inhibitory neuron lineage (*Nkx2.1*–Cre) is targeted, little effect is seen. Thus, although further studies are needed to clarify the exact nature of the affected cells, it seems that to develop FCD, pathogenic mutations must occur in progenitors of the neuronal lineage, and especially in excitatory neurons. Despite this, neurons not carrying the somatic mutation have also been shown to contribute to epilepsy in mouse models of FCD[130,131], and the contribution of cell-autonomous versus non-cell-autonomous effects to epileptogenesis is a topic of current investigation.

Autism spectrum disorder. ASD is a heterogenous disorder with a complex multigenetic cause that has not been fully clarified yet. Both common and rare de novo germline variants (CNVs, SNVs and, recently, tandem repeat mutations) have been implicated in ASD[132–136], and an excess of biallelic loss-of-function and damaging missense mutations has also been identified in approximatively 5% of cases[137]. Analyses of whole-exome sequencing data from simplex ASD families have revealed that exonic somatic mutations (SNVs and small indels) are enriched in probands compared with their unaffected siblings[138,139], and 0.8–1.3% of probands carried a missense mosaic mutation in intolerant genes potentially related to ASD risk[139]. Missense and/or loss-of-function mutations with high VAFs were estimated to increase the risk of ASD by approximatively 3.4%[138].

In addition to somatic SNVs, large CNVs (more than 4 Mb long) also give a small but significant contribution to ASD risk in 0.2% of probands, and the length of mosaic CNVs positively correlates with ASD severity[15]. Smaller mosaic CNVs (less than 4 Mb long) overlap ASD risk genes in probands more often than expected by chance, suggesting that they might also contribute to ASD risk[15]. Thus, ASD risk seems to be conferred mostly by large events that may not be tolerated at the germline state, whereas smaller events might have limited consequences at the mosaic state even when disrupting ASD-associated regions.

Most studies of somatic mutations in ASD are limited by the fact that DNAs were exclusively derived from peripheral blood, saliva or lymphoblastoid cell lines, which does not allow the detection of somatic variants restricted to brain tissue. Furthermore, whole-exome sequencing limits the analysis to transcribed regions; however, a significant contribution to ASD risk could come from intronic or intergenic variation in regulatory regions. A recent study performed deep WGS of a relatively large cohort of ASD and neurotypical brains, and suggested that ASD brains may carry an excess rate of mosaic mutations in critical brain-active enhancers associated with genes that have brain-specific expression[46].

Thus, existing data suggest that somatic SNVs and CNVs that occur early enough in development to be detectable in blood-derived DNA contribute to increasing the risk of developing ASD in as many as 5% of cases[15,139]. However, the exact proportion of somatic mutations that contribute to ASD risk compared with germline mutations is still unclear, as is what genetic backgrounds are more susceptible to develop ASD in the presence of somatic hits.

Future directions for lineage tracing

The degree of complexity and resolution of a lineage rely on (1) the number of variants confidently identified and (2) the number of single cells available for retrospective tracing. Compared with animal models, where high barcode complexity can be reached through genome editing and where genetic scars are introduced in a controlled manner, allowing targeted-sequencing, lineage tracing in humans remains extremely challenging due to the difficulty of calling low VAF mutations genome-wide. Indeed, as most of the variants useful for lineage tracing are located in intergenic regions[39], the ideal approach would be WGS of thousands of cells, which is currently prohibitively expensive. For this reason, most of the studies conducted until now have relied on variant calling in a few single cells and deep bulk WGS from a limited number of subjects, and accurate lineage reconstruction has been restricted to a few postzygotic cell generations marked by early variants. Future directions should thus point towards improving variant detection in single cells at high throughput. A method called Sci-LIANTI allows the sequencing of thousands and potentially millions of cells by combining combinatorial indexing with linear amplification of single-cell genomes[140,141]. However, sparsity of coverage of both alleles in diploid genomes is still an issue, which limits variant calling, and high coverage of the genome in thousands of cells is still prohibitively expensive. An alternative approach to reduce the costs of sequencing could be the targeted amplification and sequencing of genomic regions where developmental somatic SNVs have a higher probability to occur[39]. However, this still requires much more prior knowledge of general SNV enrichment in single neurons than we currently have. Mitochondrial single-cell assay for transposase-accessible chromatin with sequencing[142] and a similar approach called EMBLEM (epigenome and mitochondrial barcode of lineage from endogenous mutations)[143] take advantage of the small size and high somatic mutation rates of mitochondrial DNA[142]. Although this is certainly an interesting complementary method for lineage tracing, issues of mitochondrial heteroplasmy make it as yet unproven as a method of lineage tracing in complex tissues.

Another major challenge in human lineage tracing is the ability to call somatic variants in specific cell types, thus combining variant calling with cell type classification. Most studies have until now relied on the selection of specific cell populations before sequencing (for example, NeuN-positive cell sorting). However, most cell types are difficult to isolate from fresh frozen tissue specimens, and this approach is not ideal for rare cell populations and especially for diseased tissue, where expression of cell type markers may be altered. Single-cell transcriptomics and epigenomics are now the best practice to study cell types in human tissues[99,144] but the sparse coverage given by these methods significantly limits the

Epigenomics
Study of the ensemble of the epigenetic changes such as methylation and histone modifications present across the genome.

calling of somatic variants[39]. PRDD–seq has been a first method trying to overcome these challenges but it provides limited information on gene expression, and it is not ideal for diseased tissues[94]. Thus, future directions should ideally point towards developing protocols that allow DNA and RNA sequencing from the same single cell at high throughput.

Finally, adding the spatial dimension to lineage studies would allow better elucidation of the clonal relationships within the tissue architecture[145]. Although spatial transcriptomics and in situ sequencing are allowing the study of gene expression and DNA mutation genotyping in intact tissues with increasingly cellular and subcellular resolution[146], they are still limited by the low sequencing coverage, and the application of these methods to human post-mortem tissue specimens remains quite limited due to tissue quality, and still requires significant protocol optimizations.

Conclusions

Studies performed until now have shown that the human brain is a complex mosaic of intermingled early developmental lineages that seem to differ considerably in their contribution to the brain, as a result of developmental bottlenecks of stochastic clonal selections and/or different proliferation rates (FIG. 3). For later lineages derived from neural progenitors and restricted to the cortical tissue, initial characterization seems to suggest some progressive spatial restriction of nested lineages, but also clearly shows that more work is needed to understand the full extent of the complex clonal architecture of the human brain and its different regions[147]. How the huge progenitor diversity characterizing the human brain contributes to the specification of cells and especially neuronal subtypes is a long-standing question in neurodevelopment, and GM has the potential to provide important insights.

The understanding of normal lineages is the first step towards the dissection of the mechanisms of diseases associated with GM. With this in mind, future study of mosaic developmental brain disorders could include analysis of non-pathogenic clonal somatic mutations present in pathological tissue for lineage tracing, and at the same time identify those lineages affected by the pathogenic hit. In some diseases, such as FCD, mutations can alter a progenitor lineage by increasing the size of the mutant clone. However, this increase still depends on the time point in development when the somatic hit occurs and the type of progenitor cell involved. Some mutations could have a positive selection effect on some clones without necessarily leading to pathology, in some sort of non-pathological clonal expansion phenomenon. Other mutations could act in the opposite direction by negatively selecting a clone and reducing or eliminating its contribution to the adult tissue as some sort of protective mechanism. Although mutations causing clonal expansion in other normal tissues have been shown[148,149], and although this phenomenon has been suggested in the context of glial cells and clonal mutations associated with brain tumours[56], the existence and the extent of these positive and negative selection processes during human brain development and especially neurogenesis remain to be shown.

The pathogenic impact of somatic mutations depends on how many cells are affected, which can define the transition from normal to disease states. However, very little is known about the minimum MF required to cause developmental disorders such as FCD, especially due to current detection limits for bulk sequencing data. In multigenic complex disorders such as ASD, GM can contribute to pathology by adding to the effect of inherited or de novo germline mutations. Thus, the sum of several mutations that confer different degrees of risk may lead to the tolerated threshold being exceeded.

Although we are still far from understanding the full extent of clonal lineages in the human brain in the context of normal development and disease, the ability to confidently and extensively identify GM in human tissues has certainly revolutionized our approach to studying the human brain. Fresh frozen tissue specimens are now offering a unique resource for developmental studies, and retrospective lineage tracing based on GM is a promising tool for the study of neurodevelopmental disorders in humans.

Published online: 23 March 2022

1. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
2. Ludwig, L. S. et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339 (2019).
3. Spits, C. et al. Whole-genome multiple displacement amplification from single cells. *Nat. Protoc.* **1**, 1965–1970 (2006).
4. Gonzalez-Pena, V. et al. Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc. Natl Acad. Sci. USA* **118**, e2024176118 (2021).
This work presents one of the most advanced methods currently used for whole-genome amplification and genomic variant detection in single cells.
5. Xing, D., Tan, L., Chang, C. H., Li, H. & Xie, X. S. Accurate SNV detection in single cells by transposon-based whole-genome amplification of complementary strands. *Proc. Natl Acad. Sci. USA* **118**, 676–690 (2021).
6. Abascal, F. et al. Somatic mutation landscapes at single-molecule resolution. *Nature* **593**, 405–410 (2021).
7. Bohrson, C. L. et al. Linked-read analysis identifies mutations in single-cell DNA-sequencing data. *Nat. Genet.* **51**, 749–754 (2019).
8. Huang, A. Y. et al. MosaicHunter: accurate detection of postzygotic single-nucleotide mosaicism through next-generation sequencing of unpaired, trio, and paired samples. *Nucleic Acids Res.* **45**, e76 (2017).
9. Dou, Y. et al. Accurate detection of mosaic variants in sequencing data without matched controls. *Nat. Biotechnol.* **38**, 314–319 (2020).
10. Luquette, L. J., Bohrson, C. L., Sherman, M. A. & Park, P. J. Identification of somatic mutations in single cell DNA-seq using a spatial model of allelic imbalance. *Nat. Commun.* **10**, 3908 (2019).
11. Tu, K., Lu, K., Zhang, Q., Huang, W. & Xie, D. Accurate single-cell genotyping utilizing information from the local genome territory. *Nucleic Acids Res.* **49**, e57 (2021).
12. Wang, Y. et al. Comprehensive identification of somatic nucleotide variants in human brain tissue. *Genome Biol.* **22**, 92 (2021).
This work explains best practices for somatic variant detection in the human brain defined by the Brain Somatic Mosaicism Network.
13. Luquette, L. J. et al. Ultraspecific somatic SNV and indel detection in single neurons using primary template-directed amplification. Preprint at *bioRxiv* https://doi.org/10.1101/2021.04.30.442032 (2021).
14. Maury, E. A. & Walsh, C. A. Somatic copy number variants in neuropsychiatric disorders. *Curr. Opin. Genet. Dev.* **68**, 9–17 (2021).
15. Sherman, M. A. et al. Large mosaic copy number variations confer autism risk. *Nat. Neurosci.* **24**, 197–203 (2021).
16. Evrony, G. D. et al. Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).
17. Evrony, G. D., Lee, E., Park, P. J. & Walsh, C. A. Resolving rates of mutation in the brain using single-neuron genomics. *eLife* **5**, e12966 (2016).
18. Ebert, P. et al. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372**, eabf7117 (2021).
19. Mallory, X. F., Edrisi, M., Navin, N. & Nakhleh, L. Methods for copy number aberration detection from single-cell DNA-sequencing data. *Genome Biol.* **21**, 208 (2020).
20. McKerrow, W. et al. Human transposon insertion profiling by sequencing (TIPseq) to map LINE-1

insertions in single cells. *Phil. Trans. R. Soc. B* **375**, 20190335 (2020).
21. Zhu, X. et al. Machine learning reveals bilateral distribution of somatic L1 insertions in human neurons and glia. *Nat. Neurosci.* **24**, 186–196 (2021).
22. Loh, P.-R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
23. Grossmann, S. et al. Development, maturation, and maintenance of human prostate inferred from somatic mutations. *Cell Stem Cell* **7**, 1262–1274.e5 (2021).
24. Lawson, A. R. J. et al. Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science* **370**, 75–82 (2020).
25. Martincorena, I. et al. Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911–917 (2018).
26. Welch, J. S. et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
27. Osorio, F. G. et al. Somatic mutations reveal lineage relationships and age-related mutagenesis in human hematopoiesis. *Cell Rep.* **25**, 2308–2316.e4 (2018).
28. Martincorena, I. et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880–886 (2015).
29. Suda, K. et al. Clonal expansion and diversification of cancer-associated mutations in endometriosis and normal endometrium. *Cell Rep.* **24**, 1777–1789 (2018).
30. Roerink, S. F. et al. Intra-tumour diversification in colorectal cancer at the single-cell level. *Nature* **556**, 457–462 (2018).
31. Lee-Six, H. et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532–537 (2019).
32. Brunner, S. F. et al. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**, 538–542 (2019).
33. Brazhnik, K. et al. Single-cell analysis reveals different age-related somatic mutation profiles between stem and differentiated cells in human liver. *Sci. Adv.* **6**, eaax2659 (2020).
34. Lodato, M. A. et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).
This work presents rates and mechanisms of somatic variant accumulation in postmitotic single human neurons during ageing.
35. Coorens, T. H. H. et al. Inherent mosaicism and extensive mutation of human placentas. *Nature* **592**, 80–85 (2021).
36. Yoshida, K. et al. Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature* **578**, 266–272 (2020).
37. Moore, L. et al. The mutational landscape of human somatic and germline cells. *Nature* **597**, 381–386 (2021).
38. Coorens, T. H. H. et al. Extensive phylogenies of human development inferred from somatic mutations. *Nature* **597**, 387–392 (2021).
39. Bizzotto, S. et al. Landmarks of human embryonic development inscribed in somatic mutations. *Science* **371**, 1249–1253 (2021).
This work tracks lineages of human development from the first postzygotic cell generation to brain-specific clones using somatic mutations and shows asymmetries of clonal contributions to the human brain.
40. Spencer Chapman, M. et al. Lineage tracing of human development through somatic mutations. *Nature* **695**, 85–90 (2021).
41. Lodato, M. A. et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).
42. Fasching, L. et al. Early developmental asymmetries in cell lineage trees in living individuals. *Science* **371**, 1245–1248 (2021).
43. Lee-Six, H. et al. Population dynamics of normal human blood inferred from somatic mutations. *Nature* **561**, 473–478 (2018).
44. Ju, Y. S. et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* **543**, 714–718 (2017).
45. Jonsson, H. et al. Differences between germline genomes of monozygotic twins. *Nat. Genet.* **53**, 27–34 (2021).
46. Rodin, R. E. et al. The landscape of somatic mutation in cerebral cortex of autistic and neurotypical individuals revealed by ultra-deep whole-genome sequencing. *Nat. Neurosci.* **24**, 176–185 (2021).
47. Park, S. et al. Clonal dynamics in early human embryogenesis inferred from somatic mutation. *Nature* **597**, 393–397 (2021).
48. Volkova, N. V. et al. Mutational signatures are jointly shaped by DNA damage and repair. *Nat. Commun.* **11**, 2169 (2020).
49. Kiessling, A. A. et al. Genome-wide microarray evidence that 8-cell human blastomeres over-express cell cycle drivers and under-express checkpoints. *J. Assist. Reprod. Genet.* **27**, 265–276 (2010).
50. Bae, T. et al. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* **359**, 550–555 (2018).
This article presents rates and mechanisms of somatic mutation accumulation in human neuronal progenitors.
51. McConnell, M. J. et al. Mosaic copy number variation in human neurons. *Science* **342**, 632–637 (2013).
52. Knouse, K. A., Wu, J. & Amon, A. Assessment of megabase-scale somatic copy number variation using single-cell sequencing. *Genome Res.* **26**, 376–384 (2016).
53. Cai, X. et al. Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep.* **8**, 1280–1289 (2014).
54. Evrony, G. D. et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **151**, 483–496 (2012).
55. Erwin, J. A. et al. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat. Neurosci.* **19**, 1583–1591 (2016).
56. Ganz, J. et al. Rates and patterns of clonal oncogenic mutations in the normal human brain. *Cancer Discov.* **12**, 172–185 (2022).
57. Tomkova, M., Tomek, J., Kriaucionis, S. & Schuster-Bockler, B. Mutational signature distribution varies with DNA replication timing and strand asymmetry. *Genome Biol.* **19**, 129 (2018).
58. Tate, J. G. et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
59. Huang, A. Y. et al. Distinctive types of postzygotic single-nucleotide mosaicisms in healthy individuals revealed by genome-wide profiling of multiple organs. *PLoS Genet.* **14**, e1007395 (2018).
60. Wu, W. et al. Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature* **593**, 440–444 (2021).
61. Reid, D. A. et al. Incorporation of a nucleoside analog maps genome repair sites in postmitotic human neurons. *Science* **372**, 91–94 (2021).
62. Raj, B. et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
63. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018).
64. Rodriguez-Fraticelli, A. E. et al. Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216 (2018).
65. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
66. Leeper, K. et al. Lineage barcoding in mice with homing CRISPR. *Nat. Protoc.* **16**, 2088–2108 (2021).
67. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
68. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
69. Kalhor, R. et al. Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018).
This article is a very nice example of prospective lineage tracing in a model organism using genome editing to introduce cell barcodes.
70. Chan, M. M. et al. Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).
71. Llorca, A. et al. A stochastic framework of neurogenesis underlies the assembly of neocortical cytoarchitecture. *eLife* **8**, e51381 (2019).
72. Hansen, D. V., Lui, J. H., Parker, P. R. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010).
73. Xiang, L. et al. A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature* **577**, 537–542 (2020).
74. Rakic, P. Radial versus tangential migration of neuronal clones in the developing cerebral cortex. *Proc. Natl Acad. Sci. USA* **92**, 11323–11327 (1995).
75. Price, J., Turner, D. & Cepko, C. Lineage analysis in the vertebrate nervous system by retrovirus-mediated gene transfer. *Proc. Natl Acad. Sci. USA* **84**, 156–160 (1987).
76. Cepko, C. Retrovirus vectors and their applications in neurobiology. *Neuron* **1**, 345–353 (1988).
77. Huang, Z. J. & Zeng, H. Genetic approaches to neural circuits in the mouse. *Annu. Rev. Neurosci.* **36**, 183–215 (2013).
78. Joyner, A. L. & Zervas, M. Genetic inducible fate mapping in mouse: establishing genetic lineages and defining genetic neuroanatomy in the nervous system. *Dev. Dyn.* **235**, 2376–2385 (2006).
79. He, M. et al. Strategies and tools for combinatorial targeting of GABAergic neurons in mouse cerebral cortex. *Neuron* **92**, 555 (2016).
80. Livet, J. et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).
81. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
82. Zong, H., Espinosa, J. S., Su, H. H., Muzumdar, M. D. & Luo, L. Mosaic analysis with double markers in mice. *Cell* **121**, 479–492 (2005).
83. Tasic, B. et al. Extensions of MADM (mosaic analysis with double markers) in mice. *PLoS ONE* **7**, e33332 (2012).
84. Gao, P. et al. Deterministic progenitor behavior and unitary production of neurons in the neocortex. *Cell* **159**, 775–788 (2014).
85. Beattie, R. et al. Mosaic analysis with double markers reveals distinct sequential functions of Lgl1 in neural stem cells. *Neuron* **94**, 517–533.e3 (2017).
86. Bonaguidi, M. A. et al. In vivo clonal analysis reveals self-renewing and multipotent adult neural stem cell characteristics. *Cell* **145**, 1142–1155 (2011).
87. Ma, J., Shen, Z., Yu, Y. C. & Shi, S. H. Neural lineage tracing in the mammalian brain. *Curr. Opin. Neurobiol.* **50**, 7–16 (2018).
88. Reid, C. B., Tavazoie, S. F. & Walsh, C. A. Clonal dispersion and evidence for asymmetric cell division in ferret cortex. *Development* **124**, 2441–2450 (1997).
89. Ware, M. L., Tavazoie, S. F., Reid, C. B. & Walsh, C. A. Coexistence of widespread clones and large radial clones in early embryonic ferret cortex. *Cereb. Cortex* **9**, 636–645 (1999).
90. Kornack, D. R. & Rakic, P. Radial and horizontal deployment of clonally related cells in the primate neocortex: relationship to distinct mitotic lineages. *Neuron* **15**, 311–321 (1995).
91. Betizeau, M. et al. Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442–457 (2013).
92. Gertz, C. C., Lui, J. H., LaMonica, B. E., Wang, X. & Kriegstein, A. R. Diverse behaviors of outer radial glia in developing ferret and human cortex. *J. Neurosci.* **34**, 2559–2570 (2014).
93. Lin, Y. et al. Behavior and lineage progression of neural progenitors in the mammalian cortex. *Curr. Opin. Neurobiol.* **66**, 144–157 (2021).
This is a nice recent review of what is known about mammalian cortical progenitors and their lineage output, including differences between lissencephalic and gyrencephalic species such as humans.
94. Huang, A. Y. et al. Parallel RNA and DNA analysis after deep sequencing (PRDD-seq) reveals cell type-specific lineage patterns in human brain. *Proc. Natl Acad. Sci. USA* **117**, 13886–13895 (2020).
This work presents a recently developed method to detect clonal somatic variants and expression of cell type-specific gene markers in the same single cell to couple lineage tracing with cell type information in the human brain.
95. Delgado, R. N. et al. Individual human cortical progenitors can produce excitatory and inhibitory neurons. *Nature* **601**, 397–403 (2021).
96. Hansen, D. V. et al. Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat. Neurosci.* **16**, 1576–1587 (2013).
97. Rudy, B., Fishell, G., Lee, S. & Hjerling-Leffler, J. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).
98. Nigro, M. J., Hashikawa-Yamasaki, Y. & Rudy, B. Diversity and connectivity of layer 5 somatostatin-expressing interneurons in the mouse barrel cortex. *J. Neurosci.* **38**, 1622–1633 (2018).

99. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
This work presents one of the most updated classifications of human brain cell types obtained from single-cell transcriptomics.
100. Ang, E. S. Jr., Haydar, T. F., Gluncic, V. & Rakic, P. Four-dimensional migratory coordinates of GABAergic interneurons in the developing mouse cortex. *J. Neurosci.* **23**, 5805–5815 (2003).
101. Rymar, V. V. & Sadikot, A. F. Laminar fate of cortical GABAergic interneurons is dependent on both birthdate and phenotype. *J. Comp. Neurol.* **501**, 369–380 (2007).
102. Nemtsova, M. V. et al. Clinical relevance of somatic mutations in main driver genes detected in gastric cancer patients by next-generation DNA sequencing. *Sci. Rep.* **10**, 504 (2020).
103. van Rooij, J. et al. Somatic TARDBP variants as a cause of semantic dementia. *Brain* **143**, 3827–3841 (2020).
104. Mass, E. et al. A somatic mutation in erythro-myeloid progenitors causes neurodegenerative disease. *Nature* **549**, 389–393 (2017).
105. 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 (2019).
106. Lobon, I. et al. Somatic mutations in Parkinson disease are enriched in synaptic and neuronal processes. Preprint at *medRxiv* https://doi.org/10.1101/2020.09.14.20190538 (2020).
107. Miller, M. B., Reed, H. C. & Walsh, C. A. Brain somatic mutation in aging and Alzheimer's disease. *Annu. Rev. Genomics Hum. Genet.* **22**, 239–256 (2021).
108. Doan, R. N. et al. Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341–354.e12 (2016).
109. Lim, E. T. et al. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nat. Neurosci.* **20**, 1217–1224 (2017).
110. Kryukov, G. V., Pennacchio, L. A. & Sunyaev, S. R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.* **80**, 727–739 (2007).
111. Poduri, A. et al. Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron* **74**, 41–48 (2012).
112. Conti, V. et al. Focal dysplasia of the cerebral cortex and infantile spasms associated with somatic 1q21.1-q44 duplication including the AKT3 gene. *Clin. Genet.* **88**, 241–247 (2015).
113. Kobow, K. et al. Mosaic trisomy of chromosome 1q in human brain tissue associates with unilateral polymicrogyria, very early-onset focal epilepsy, and severe developmental delay. *Acta Neuropathol.* **140**, 881–891 (2020).
114. Kim, M.-H. et al. Low-level brain somatic mutations are implicated in schizophrenia. *Biol. Psychiatry* **90**, 35–46 (2021).
115. Fullard, J. F. et al. Assessment of somatic single-nucleotide variation in brain tissue of cases with schizophrenia. *Transl. Psychiatry* **9**, 21 (2019).
116. Blümcke, I. et al. Toward a better definition of focal cortical dysplasia: an iterative histopathological and genetic agreement trial. *Epilepsia* **62**, 1416–1428 (2021).
117. D'Gama, A. M. et al. Somatic mutations activating the mTOR pathway in dorsal telencephalic progenitors cause a continuum of cortical dysplasias. *Cell Rep.* **21**, 3754–3766 (2017).
118. Jansen, L. A. et al. PI3K/AKT pathway mutations cause a spectrum of brain malformations from megalencephaly to focal cortical dysplasia. *Brain* **138**, 1613–1628 (2015).
119. Lim, J. S. et al. Brain somatic mutations in MTOR cause focal cortical dysplasia type II leading to intractable epilepsy. *Nat. Med.* **21**, 395–400 (2015).
120. Lee, W. S. et al. Gradient of brain mosaic RHEB variants causes a continuum of cortical dysplasia. *Ann. Clin. Transl. Neurol.* **8**, 485–490 (2021).
121. Baldassari, S. et al. Dissecting the genetic basis of focal cortical dysplasia: a large cohort study. *Acta Neuropathol.* **138**, 885–900 (2019).
This article shows the role of pathogenic brain somatic mutations in a wide cohort of patients affected by epileptic FCD.
122. Lee, J. H. et al. De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat. Genet.* **44**, 941–945 (2012).
123. Lim, J. S. et al. Somatic mutations in TSC1 and TSC2 cause focal cortical dysplasia. *Am. J. Hum. Genet.* **100**, 454–472 (2017).
124. Ribierre, T. et al. Second-hit mosaic mutation in mTORC1 repressor DEPDC5 causes focal cortical dysplasia-associated epilepsy. *J. Clin. Invest.* **128**, 2452–2458 (2018).
125. Winawer, M. R. et al. Somatic SLC35A2 variants in the brain are associated with intractable neocortical epilepsy. *Ann. Neurol.* **83**, 1133–1146 (2018).
126. Bonduelle, T. et al. Frequent SLC35A2 brain mosaicism in mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE). *Acta Neuropathol. Commun.* **9**, 3 (2021).
127. Blumcke, I. et al. The clinicopathologic spectrum of focal cortical dysplasias: a consensus classification proposed by an ad hoc task force of the ILAE Diagnostic Methods Commission. *Epilepsia* **52**, 158–174 (2011).
128. Cepeda, C. et al. Epileptogenesis in pediatric cortical dysplasia: the dysmature cerebral developmental hypothesis. *Epilepsy Behav.* **9**, 219–235 (2006).
129. Koh, H. Y. et al. BRAF somatic mutation contributes to intrinsic epileptogenicity in pediatric brain tumors. *Nat. Med.* **24**, 1662–1668 (2018).
130. Koh, H. Y. et al. Non-cell autonomous epileptogenesis in focal cortical dysplasia. *Ann. Neurol.* **90**, 285–299 (2021).
This article presents a mouse model of FCD due to brain mTOR activating somatic mutations and nicely explains the cell-autonomous versus non-cell-autonomous effects in mosaic brain disorders.
131. Nguyen, L. H. et al. Genetic expression of 4E-BP1 in juvenile mice alleviates mTOR-induced neuronal dysfunction and epilepsy. *Brain* https://doi.org/10.1093/brain/awab390 (2021).
132. Weiner, D. J. et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* **49**, 978–985 (2017).
133. Krumm, N. et al. Excess of rare, inherited truncating mutations in autism. *Nat. Genet.* **47**, 582–588 (2015).
134. Sanders, S. J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
135. Mitra, I. et al. Patterns of de novo tandem repeat mutations and their role in autism. *Nature* **589**, 246–250 (2021).
136. Iossifov, I. et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
137. Doan, R. N. et al. Recessive gene disruptions in autism spectrum disorder. *Nat. Genet.* **51**, 1092–1098 (2019).
138. Dou, Y. et al. Postzygotic single-nucleotide mosaicisms contribute to the etiology of autism spectrum disorder and autistic traits and the origin of mutations. *Hum. Mutat.* **38**, 1002–1013 (2017).
This work analyses whole-exome sequencing data to show the contribution of somatic mosaic mutations to ASDs.
139. Krupp, D. R. et al. Exonic mosaic mutations contribute risk for autism spectrum disorder. *Am. J. Hum. Genet.* **101**, 369–390 (2017).
140. Yin, Y. et al. High-throughput single-cell sequencing with linear amplification. *Mol. Cell* **76**, 676–690.e10 (2019).
This article presents one of the most high-throughput single-cell sequencing methods now available.
141. Chen, C. et al. Single-cell whole-genome analyses by linear amplification via transposon insertion (LIANTI). *Science* **356**, 189–194 (2017).
142. Lareau, C. A. et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. *Nat. Biotechnol.* **39**, 451–461 (2020).
143. Xu, J. et al. Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. *eLife* **8**, e45105 (2019).
144. Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
145. Chow, K.-H. K. et al. Imaging cell lineage with a synthetic digital recording system. *Science* **372**, eabb3099 (2021).
146. Liao, J., Lu, X., Shao, X., Zhu, L. & Fan, X. Uncovering an organ's molecular architecture at single-cell resolution by spatially resolved transcriptomics. *Trends Biotechnol.* **39**, 43–58 (2021).
147. Pollen, A. A. et al. Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).
148. Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **173**, 1823 (2018).
149. Kakiuchi, N. & Ogawa, S. Clonal expansion in non-cancer tissues. *Nat. Rev. Cancer* **21**, 239–256 (2021).

Acknowledgements
The authors thank K. Probst (Xavier Studio) for valuable assistance with figure assembly and design. They also thank all their colleagues at Boston Children's Hospital and Harvard Medical School for discussions that helped formulate and clarify ideas included in this Review. C.A.W. is an investigator of the Howard Hughes Medical Institute, supported by the Allen Frontiers Group through the Allen Discovery Center for Human Brain Evolution and by NINDS grants R01NS32457 and R01NS35129. S.B. was supported by the Manton Center for Orphan Disease Research at Boston Children's Hospital and is now supported by a European Commission's Horizon 2020 Research and Innovation Programme Marie Skłodowska-Curie Actions Individual Fellowship (grant agreement no. 101026484 — CODICES).

Author contributions
S.B. and C.A.W. wrote the Review and contributed substantially to discussion of the content.

Competing interests
C.A.W. is a paid consultant (cash and no equity) for Third Rock Ventures and is on the Clinical Advisory Board (cash and equity) of Maze Therapeutics. No research support was received. These companies did not fund and had no role in the writing of this Review. S.B. declares no competing interests.

Peer review information
Nature Reviews Neuroscience thanks B. Treutlein, J.H. Lee and the other, anonymous reviewers for their contribution to the peer review of this work.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.