Distinctive types of postzygotic single-nucleotide mosaicisms in healthy individuals revealed by genome-wide profiling of multiple organs

August Yue Huang[1]☉*, Xiaoxu Yang[1]☉, Sheng Wang[2,3], Xianing Zheng[2], Qixi Wu[4,5], Adam Yongxin Ye[1,4,6], Liping Wei[1]*

1 Center for Bioinformatics, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, China, 2 National Institute of Biological Sciences, Beijing, China, 3 College of Biological Sciences, China Agricultural University, Beijing, China, 4 Peking-Tsinghua Center for Life Sciences, Beijing, China, 5 Human Genetics Resources Core Facility, School of Life Sciences, Peking University, Beijing, China, 6 Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China

☉ These authors contributed equally to this work.
* weilp@mail.cbi.pku.edu.cn (LW); huangy@mail.cbi.pku.edu.cn (AYH)

Abstract

Postzygotic single-nucleotide mosaicisms (pSNMs) have been extensively studied in tumors and are known to play critical roles in tumorigenesis. However, the patterns and origin of pSNMs in normal organs of healthy humans remain largely unknown. Using whole-genome sequencing and ultra-deep amplicon re-sequencing, we identified and validated 164 pSNMs from 27 postmortem organ samples obtained from five healthy donors. The mutant allele fractions ranged from 1.0% to 29.7%. Inter- and intra-organ comparison revealed two distinctive types of pSNMs, with about half originating during early embryogenesis (embryonic pSNMs) and the remaining more likely to result from clonal expansion events that had occurred more recently (clonal expansion pSNMs). Compared to clonal expansion pSNMs, embryonic pSNMs had higher proportion of C>T mutations with elevated mutation rate at CpG sites. We observed differences in replication timing between these two types of pSNMs, with embryonic and clonal expansion pSNMs enriched in early- and late-replicating regions, respectively. An increased number of embryonic pSNMs were located in open chromatin states and topologically associating domains that transcribed embryonically. Our findings provide new insights into the origin and spatial distribution of postzygotic mosaicism during normal human development.

Author summary

Genomic mosaicism led by postzygotic mutation is the major cause of cancers and many non-cancer developmental disorders. Theoretically, postzygotic mutations should be accumulated during the developmental process of healthy individuals, but the genome-

wide characterization of postzygotic mosaicisms across many organ types of the same individual remained limited. In this study, we identified and validated two types of postzygotic mosaicism from the whole-genomes of 27 organs obtained from five healthy donors. We further found that the postzygotic mosaicisms arising during early embryogenesis and later clonal expansion events show distinct genomic patterns in mutation spectrum, replication timing, and chromatin status.

Introduction

Postzygotic mutations refer to DNA changes arising after the formation of the zygote that lead to genomic mosaicisms in a single individual [1, 2]. Unlike *de novo* or inherited germline mutations, postzygotic mutations only affect a fraction of cells in multicellular organisms, and individuals carrying a functional mosaic mutation typically exhibit a milder phenotype [3–5]. The roles of postzygotic single-nucleotide mosaicisms (pSNMs) have been demonstrated in numerous cancers [6, 7] and various types of developmental disorders, including malformations [8, 9] and autism [10, 11]. We and another research group have reported the first genome-wide identification and characterization of pSNMs from the peripheral blood samples of healthy individuals [12, 13]. More recently, the accumulation of postzygotic mutations during aging process has been reported in blood or brain samples [14–17]. Yadav *et al.* studied pSNMs in apparently benign tissue samples obtained from cancer patients [18], but the contribution of pre-cancerous mutations could not be completely ruled out and the study was restricted to exonic regions. As such, the occurrence and genomic pattern of pSNMs in normal tissues of healthy individuals remains under-investigated.

It has been reported that cancer genomes have distinct mutational signatures resulting predominantly from exposure to mutagenic agents and dysfunction of the DNA repair machinery [19]. Additional genomic factors, such as replication timing and chromatin status, could also impact the distribution of pSNMs in cancer genomes [20–22]. Whether and how these genomic factors might contribute to the genomic distribution of pSNMs in organs of healthy individuals remains largely unexplored [23]. Tumorigenesis has been considered as an evolutionary process in which tumor cells with increased fitness will proliferate faster than normal cells and lead to the clonal expansion of tumor cell population in a specific organ [24, 25]. Although such events of clonal expansion have been previously reported in apparently normal skin and blood samples [17, 26], it remains unclear whether clonal expansion plays a role in other non-cancer tissue types. Understanding the origin and spatial distribution of pSNMs in normal tissues of healthy individuals could provide an important baseline for interpreting their contributions to disease states [27].

The next-generation sequencing technologies (NGS) have greatly advanced the study of pSNMs [28]. Sequencing the genomes of single cells after whole-genome amplification or *in vivo* clonal proliferation have been applied to the study of pSNM profiles of normal human cells, including germ cells [29], adult stem cells [30], and neurons [31]. Typically, tens or hundreds of cells from each sample need to be sequenced to identify and quantify pSNMs, which tends to increase the cost [32]. The inaccurate process of whole-genome amplification in single-cell sequencing makes it difficult to distinguish real pSNMs from technical artifacts, and the challenge of rigorously validating the pSNMs in a cell that have been already amplified aggravates the uncertainties [33, 34]. Bulk sequencing is potentially a reliable and cost-effective alternative that, importantly, allows for rigorous validations of pSNMs [23]. Utilizing bulk whole-genome sequencing (WGS) and ultra-deep amplicon re-sequencing, this current study

identified and validated pSNMs from 27 different organ samples obtained from five healthy donors and investigated the origin and spatial distribution of pSNMs in the developmental process of these healthy individuals.

Results

Genome-wide identification and profiling of postzygotic single-nucleotide mosaicisms in healthy human organ samples

Postmortem organ samples derived from five healthy Asian donors (age 20–45 yr) were obtained from BioServe, including a total of 27 organ samples from brain, liver, colon, skin, artery, breast, ovary, and prostate (Table 1). The donors died from motor vehicle accidents and were not known to be affected by any types of cancer or other overgrowth disorders. The samples were sequenced using an Illumina HiSeq X Ten sequencing platform with an average depth of 114-150X (Table 1).

It is expected that many postzygotic mutations occurring at an early stage of embryogenesis may be shared between two or more organs from one individual [13, 27]. Thus, conventional mutation callers, which require matched negative control samples for comparison, would likely miss these mutations. We had previously developed MosaicHunter [35], a bioinformatics

Table 1. Donor and sequencing information regarding the 27 post-mortem organ samples profiled in this study.

Individual	Race	Gender	Age	Organ	WGS Depth		Number of pSNMs	
					Raw	Final	Candidate	Validated
BBL1100C	Asian	Male	40	Brain	122.53	90.88	10	25	16
				Liver	121.05	90.49	9		
				Skin	114.37	83.34	10		
				Colon	117.80	88.44	4		
				Prostate	121.69	88.36	8		
BBL11121	Asian	Female	45	Brain	141.53	109.06	10	78	59
				Liver	140.97	108.41	10		
				Skin	145.13	107.24	20		
				Colon	144.93	107.62	11		
				Breast	145.47	113.44	45		
				Ovary	136.86	106.38	6		
BBLC1013	Asian	Male	20	Brain	139.01	105.93	12	39	16
				Liver	125.57	94.19	17		
				Skin	130.52	99.09	8		
				Colon	124.77	94.24	10		
				Artery	136.29	104.63	8		
				Prostate	133.51	102.26	11		
BBLD1005	Asian	Male	29	Brain	124.86	92.36	9	77	54
				Liver	134.09	105.00	60		
				Skin	118.10	87.44	6		
				Colon	128.93	100.51	8		
				Artery	130.35	100.51	8		
BBLD1010	Asian	Male	40	Brain	129.49	96.43	9	32	19
				Liver	127.44	96.83	8		
				Skin	137.66	101.37	10		
				Colon	149.94	88.27	7		
				Prostate	130.35	99.62	10		

https://doi.org/10.1371/journal.pgen.1007395.t001

pipeline that can detect pSNMs without the need of control sample obtained from the same individual. MosaicHunter incorporated a Bayesian genotyper to distinguish pSNMs from germline variants and base-calling errors and a series of stringent filters to remove systematic errors. Using the Bayesian genotyper, we calculated the posterior probability of mosaic genotype versus three germline genotypes across all the genomic sites with at least 5% mutant allele fraction (the fraction of reads supporting the mutant allele) and 3 or more reads supporting the mutant allele. As a result, we identified a total of 251 candidate pSNMs in the 27 samples from the five donors; among them 41 pSNMs were found in more than one sample from the same donor (Table 1).

Next, we validated the pSNMs and quantified their minor allele fractions in all of the organ samples (Methods). We used an amplicon-based ultra-deep resequencing method, PASM (PGM Amplicon Sequencing of Mosaicism), which we had previously developed and benchmarked [5]. Of the 251 candidate pSNM sites, 27 were excluded due to failure to design amplicon primers or to get enough sequencing depth in the negative controls. For the remaining 224 sites, the average sequencing depth of the amplicons was greater than 4000X per sample (S1 Fig). The peripheral blood samples of two unrelated healthy Asians (ACC1 and ACC4) served as negative controls. A pSNM was considered validated only if the mutant allele was detected in an organ sample in a mosaic state but undetectable in both the negative controls. Three sites with abnormal copy numbers estimated from the WGS data were further excluded (S1 Table and Methods). In summary, we successfully validated 164 pSNMs in these five donors, with an overall validation rate of 73.2% (Table 1). The full list of the 164 validated pSNMs was described in S2 Table, which was used in the following analyses.

The validated pSNMs were located in 21 autosomes and the X chromosome (Fig 1A). We calculated the genomic distance between nearby pSNMs and found no significant difference between the observed and expected distances if pSNMs were uniformly distributed along the human genome (Kolmogorov-Smirnov test, P-value > 0.05), indicating that there was no
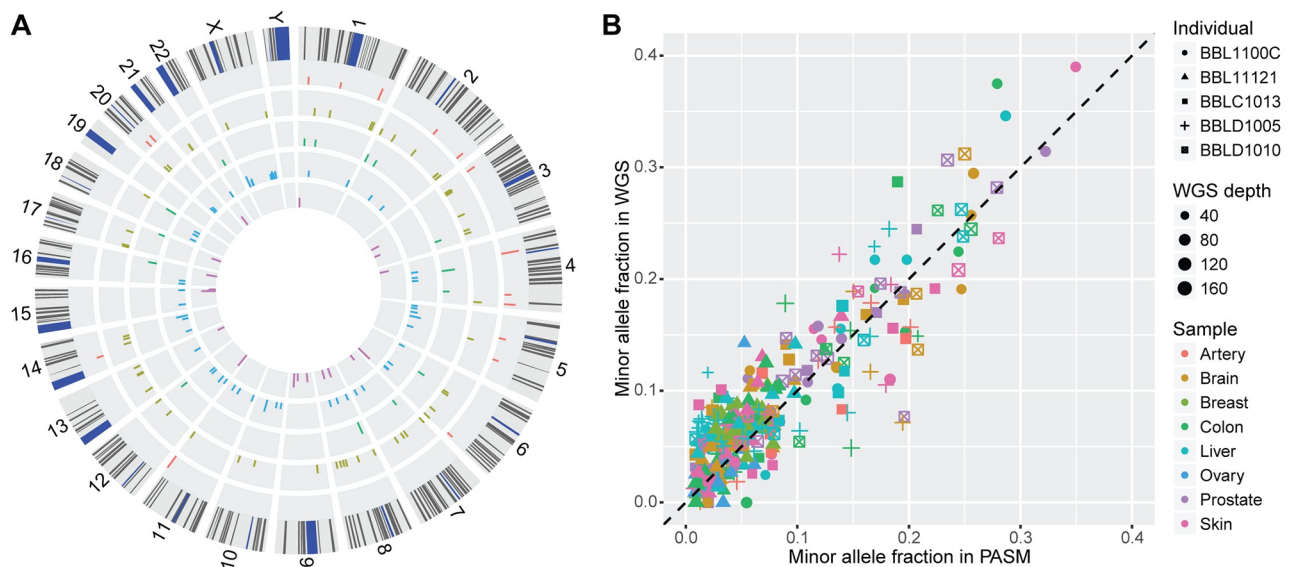

Fig 1. Identification and validation of pSNMs in 27 organ samples obtained from five individuals. (A) Genomic landscape of the validated pSNMs. Circos plots from outer to inner represent individuals BBL1100C, BBL11121, BBLC1013, BBLD1005, and BBLD1010, respectively. The Y axis denotes the average minor allele fraction across pSNM-carrying organs of each donor. (B) Correlation of the allele fractions estimated by whole-genome sequencing and targeted ultra-deep resequencing (PASM) of the validated sites. The shape, color, and size of the dots represent the donor and organ type carrying the pSNMs and the site-specific depth of whole-genome sequencing.

https://doi.org/10.1371/journal.pgen.1007395.g001

observable clustering of postzygotic mutations in healthy individuals. The minor allele fractions estimated by PASM ranged from 1.0% to 29.7%, significantly correlating with the fractions estimated by WGS (Fig 1B; Pearson's r = 0.89 and P-value < 2.2×10^{-16}). The allele fraction of each pSNM varied across the different organs from the same donor (S2–S6 Figs).

Two distinct types of postzygotic single-nucleotide mosaicisms revealed by inter- and intra-organ comparisons

Based on the presence or absence of the validated pSNMs in the organ samples from an individual donor, we grouped the pSNMs into two categories: 60 were present in two or more organ samples of the same donor (27 of which were globally present in all the sequenced organs of the donor) and 104 were uniquely present in a single organ. Given the low postzygotic mutation rate in healthy individuals [36], it was unlikely that multiple postzygotic mutation events involving the same nucleotide alteration occurred independently within one individual. It was more likely that the pSNMs shared by more than one organ resulted from mutation events that had occurred at early developmental stages, and the mutant alleles were passed on to cell lineages of more than one organ type. Comparison of the minor allele fractions in the two categories of pSNMs supported this hypothesis. As shown in Fig 2A, the allele fraction of pSNMs shared by more than one organ was significantly higher than that of pSNMs unique to only one organ (Wilcoxon rank-sum test, P-value = 1.2×10^{-3}). In particular, 40% (24 out of 60) of the pSNMs shared by more than one organ had allele fractions greater than 1/16, suggesting that they might have originated during the first few cell divisions of embryogenesis [37]. We refer to these pSNMs shared by more than one organ as "embryonic pSNMs" in the following analyses. On average, we identified 4.6~14.5 embryonic pSNMs from each organ of the five individuals, and the occurrence rate was similar across different organs (Fig 2B). We further compared the allele fractions across multiple organs of the same individual, and found that more than 95% of the embryonic pSNMs showed <5% standard deviation of allele fraction (S2 Table), indicating no dramatic allele fraction change for embryonic pSNMs.

Close inspection of the pSNMs unique to only one organ revealed a distinctive type of pSNMs. Two organs had a dramatic excess of organ-unique pSNMs compared to other organs (Fig 2B). Specifically, the liver sample of BBLD1005 and the breast sample of BBL11121 carried 42 and 32 organ-unique pSNMs, respectively, compared to an average of 1.1 organ-unique pSNMs for the other organ samples. This suggested that the majority of pSNMs in these two organs might originate organ-specifically after embryogenesis [30].

To further investigate these excessive organ-unique pSNMs in these two organs, we sampled three additional adjacent samples from each organ with varying physical distances to the original samples used for WGS (S7 Fig) and applied PASM to profile the allele fractions of validated pSNMs (S8 and S9 Figs). While 8 of the 9 (88.9%) embryonic pSNMs could be detected in all three intra-organ samples (Fig 2C and 2D), consistent with our prediction that these mutations occurred early in embryogenesis, the organ-unique pSNMs manifested with a distinct intra-organ pattern. In BBLD1005, 22 out of 42 (52%) liver-unique pSNMs identified in the original sample (liver #9) were also detected in the physically closest sample (liver #8), whereas only one liver-unique pSNM was detected in the two samples further away (liver #2 and #5) (Fig 2C). Given that the physical distance between liver #8 and #9 was about 0.5 cm and the distance between liver #2/#5 and liver #9 was approximately 3.5 and 2 cm, respectively, these results suggested that the majority of liver-unique pSNMs were locally restricted to a small volume of liver cells. A similar observation was made in the breast samples of BBL11121 that breast #7 shared more pSNMs to breast #9 than breast #2 and #5 (Fig 2D). We reconstructed the inter-sample similarity using the minor allele fractions of pSNMs, and indeed the
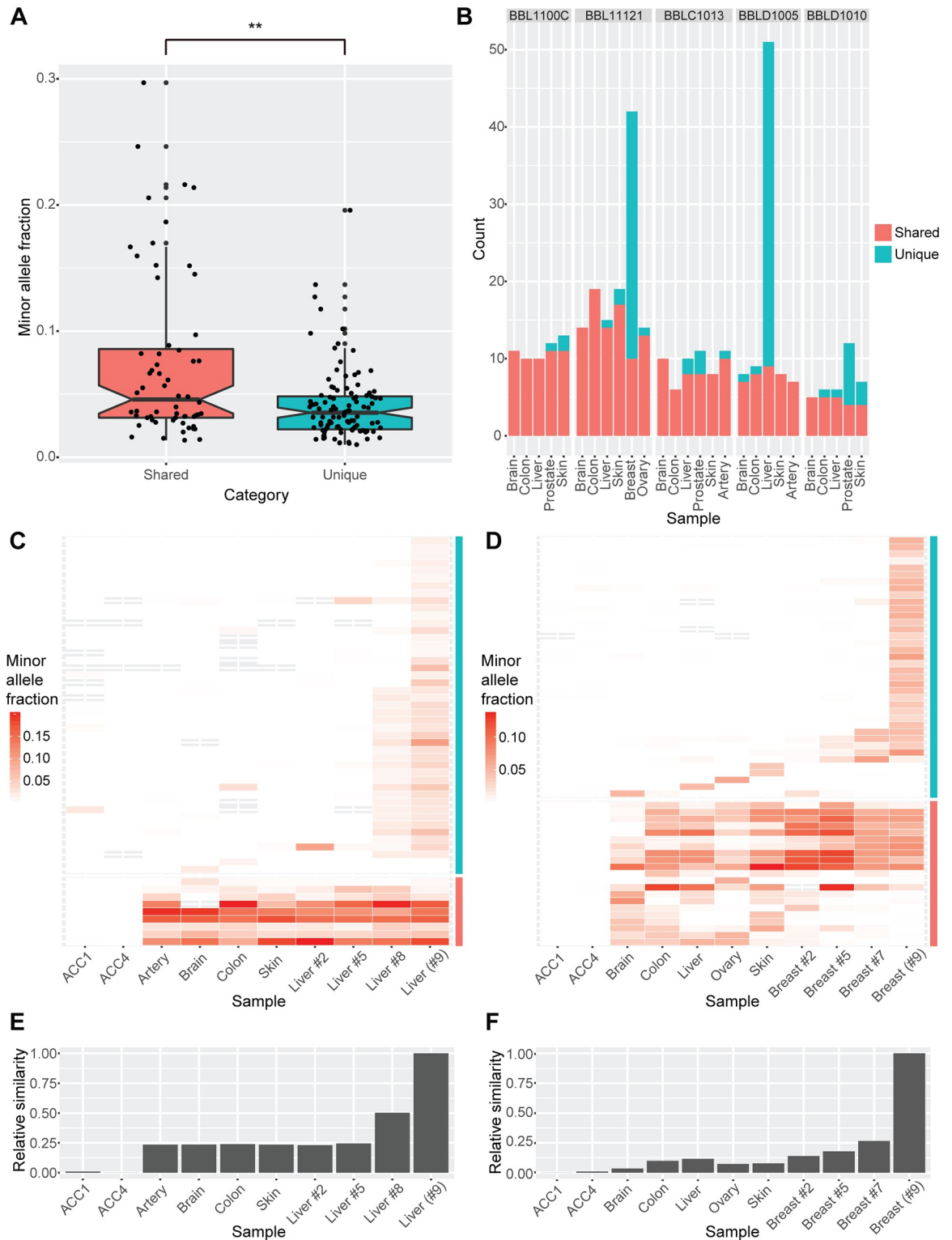
Fig 2. Two types of pSNMs revealed by inter- and intra-organ profiles. (A) Minor allele fractions between different categories of pSNM. The organ-shared pSNMs demonstrated significantly higher allele fractions than the organ-unique pSNMs. (B) Number of pSNMs carried in different organ samples. Red and blue bars denote the organ-shared and organ-unique pSNMs, respectively. An excess of organ-unique pSNMs was observed in the breast sample of BBL11121 and the liver sample of BBLD1005. (C-D) Heatmap of minor allele fractions for pSNMs carried in multiple organ samples of BBLD1005 (C) and BBL11121 (D). Blood samples of two unrelated

originally-sequenced liver or breast samples shared the largest similarity to their physically nearest samples (Fig 2E and 2F).

Analysis of minor allele fractions of the liver- and breast-unique pSNMs revealed a single narrow peak for each organ sample (S10 Fig), with an average of 3.1% and 4.2%, respectively. Considering that such pSNMs were restricted to a small region within the organ, the narrow peaks likely resulted from clonal expansion events during the process of organ self-renewal that generated a sub-population of cells carrying postzygotic mutations large enough to be detected in bulk sequencing [23]. We refer to these pSNMs as "clonal expansion pSNMs" in the following analyses. Our results demonstrated the presence of clonal expansions in various types of non-cancer organs and highlighted clonal expansion as one of the major sources of pSNMs in clinically unremarkable individuals.

Mutation spectrum, replication timing, and chromatin status varied between the two types of postzygotic single-nucleotide mosaicisms

If the embryonic pSNMs arose from early mutations during embryogenesis and the clonal expansion pSNMs arose from more recent mutations during organ self-renewal, they may present different mutational characteristics. To explore this possibility, we compared these mutations in terms of mutation spectrum, replication timing, and chromatin status.

We first studied the mutation spectrum of the two types of pSNMs identified. For embryonic pSNMs, C>T mutations were the most predominant type (65.0%), with a significant elevated mutation rate at CpG sites vs non-CpG sites (Proportion Z-test, P-value $< 2.2 \times 10^{-16}$, Fig 3A). The enrichment of C>T mutation at CpG sites could be explained by the spontaneous deamination of 5-methylcytosines (5mC) [20], which has also been reported as one of the most common signatures in cancers [38]. The predominant C>T mutation at CpG sites for embryonic pSNMs were consistent with previous studies of early pSNMs in human [14, 39] and mouse [40]. On the contrary, we observed predominant C>A (39.5%) and T>C (42.4%) mutations for the clonal expansion pSNMs identified in BBLD1005's liver and BBL11121's breast samples, respectively (Fig 3B and 3C). Oxidative DNA damage was one of the major cause for C>A mutations [22], and the higher proportion of C>A mutation in the liver sample could be explained by the accumulated oxidative stress of hepatocytes.

Previous studies had reported elevated rates of germline and cancer-related somatic mutations in late-replicating regions [41, 42]. Using data from the replication timing profile of lymphoblastoid cell-lines [43], we observed significantly different distributions of replication timing between embryonic and clonal expansion pSNMs (Wilcoxon rank-sum test, P-value = 9.7×10^{-3}; Fig 3D). Clonal expansion pSNMs were significantly enriched in late-replication regions (Permutation test, P-value = 0.006), similar to previous reports of germline and cancer-related somatic mutations, while embryonic pSNMs were significantly enriched in genomic regions that replicated earlier (Permutation test, P-value = 0.026). Embryonic pSNMs with a wide range of allele fractions contributed to the early-replication enrichment (S11 Fig), suggesting that the enrichment was not caused by a small number of outliers. This bimodal distribution was confirmed using the replication timing profiles from five other cell-lines: GM12878, K562, HeLa-S1, HepG2, and HUVEC (Wilcoxon rank-sum test, P-value < 0.05).
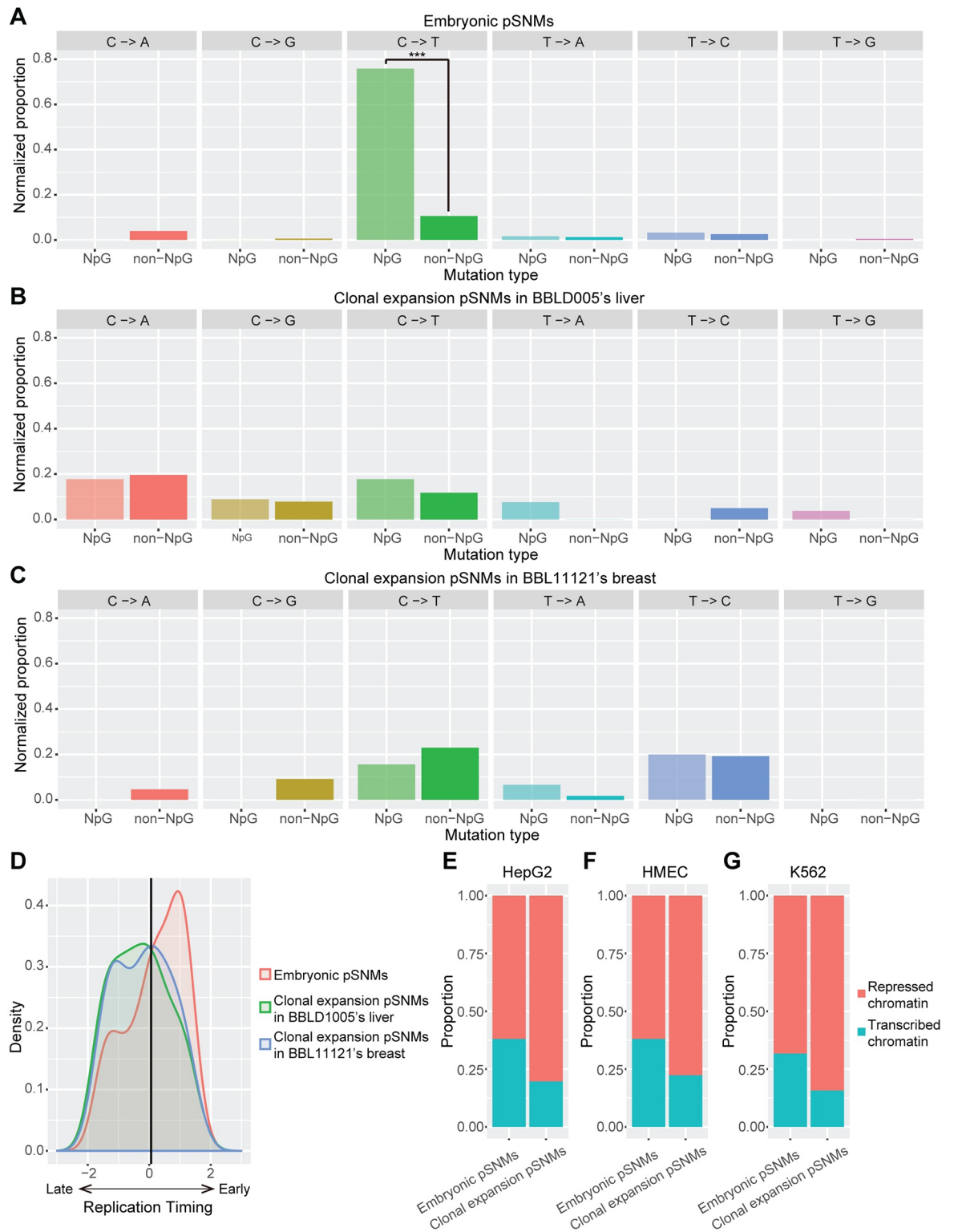
Fig 3. Distinct genomic characteristics between embryonic and clonal expansion pSNMs. (A-C) Mutation spectrums in NpG and non-NpG sites for embryonic pSNMs (A), clonal expansion pSNMs in BBLD1005's liver (B), and clonal expansion pSNMs in BBL11121's breast (C). Mutation rate was normalized by the total number of sites in the human genome. For embryonic pSNMs, CpG sites showed significantly higher rate of C>T mutations than non-CpG sites. (D) Varied DNA replication timing of the two pSNMs types. The grey line denotes the genomic average. Embryonic pSNMs were enriched in early-replicating regions, whereas clonal expansion pSNMs were

enriched in late-replicating regions. (E-G) Proportion of pSNMs locating in open or closed chromatin regions in the HepG2 (E), HMEC (F), and K562 (G) cell-lines. Significantly higher proportions of embryonic pSNMs were observed in transcribed chromatin regions of all three cell types.

We further confirmed our finding by using the pSNMs identified from the single-clone sequencing of neuronal progenitor cells [14], where the mutations which were shared by other brain regions and non-brain tissues were significantly enriched in early-replicating regions than those specifically present in the clone of neuronal progenitor cells (Wilcoxon rank-sum test, P-value = 0.028). The distinct pattern of replication timing between the two types of pSNMs might reflect different mutational effects of replication timing during different stages of human development.

Last but not the least, we investigated whether chromatin status contributed to the mutation rate of pSNMs. For embryonic pSNMs, the genomic distance between a pSNM and its closest DNase sensitive zone in embryonic stem cells was significantly smaller than the expectation under uniform distribution (Permutation test, P-value = 0.013). In contrast, clonal expansion pSNMs did not showed the enrichment of DNase sensitive zone (Permutation test, P-value = 0.54). We further found that embryonic pSNMs were significantly enriched in the topologically associating domains (TADs) containing embryonically-transcribed genes (Fisher's exact test, P-value = 0.046), and this pattern was robust with different thresholds for embryonically-transcribed genes (S12 Fig). Moreover, we observed a significantly larger proportion of embryonic pSNMs compared to clonal expansion pSNMs within transcribed chromatin regions using epigenetic data from three cell-lines of different origins (Fisher's exact test, P-value < 0.05; Fig 3E–3G). Analyses of tissue-shared pSNMs versus clone-specific pSNMs previously identified in neuronal progenitor cells [14] further confirmed our finding (Fisher's exact test, P-value < 0.01; S13 Fig). In summary, we reported an elevated rate of postzygotic mutations in open and transcribed chromatin regions during embryogenesis, which might result from the exposure of external or internal mutagens within these regions [44].

Discussion

Recent researches have significantly expanded what is known about the functional roles of postzygotic mutations, which now include not only cancers and overgrowth disorders [45], but also other complex disorders [11, 46]. With the help of next-sequencing technologies, postzygotic mutations have been identified and validated in healthy individuals [13, 39], confirming the theoretical predictions that postzygotic mutations are prevalent and every person is a mosaic [23]. However, the number of rigorously validated postzygotic mutations in healthy individuals has been small, which has hindered statistical analyses of their genomic patterns. In particular, little is known about the genomic patterns of postzygotic mutations in the normal development process of healthy human organs.

In this study, we discovered two distinct types of pSNMs, one occurring during early embryogenesis and the other likely to occur during later tissue-specific clonal expansion. Surprisingly, these mutations manifested many distinct features in regard to mutation spectrum, replication timing, and chromatin status, implying dynamic mutational effects across different developmental stages. Unsurprising in hind sight, clonal expansion pSNMs shared many mutational features with previously reported cancer mutations [47], as tumorigenesis is a specialized process involving clonal expansion of cancer cells [28]. Previous studies reported high proportion of C>A and T>C mutations as well as enrichment of late-replicating regions for clonal expansion pSNMs that were identified from skin fibroblasts [48–50], which was

concordant with our findings of clonal expansion pSNMs in the liver and breast samples. In contrast, our embryonic pSNMs demonstrated a range of unique features, including an elevated C>T mutation rate in CpG sites, an enrichment in early-replicating regions, and a stronger effect of transcribed chromatin status (Fig 3). Similar patterns in mutation spectrum (S14 Fig), replication timing (S15 Fig), and chromatin status (S3 Table) could be observed between the embryonic pSNMs that were globally present in all the sequenced organs of the donor and those only present in some but not all the sequenced organs. To further cross-validate our findings, we further analyzed an independent pSNM list that had been identified from human neuronal progenitor cells [14], and confirmed the varied genomic patterns between pSNMs which originated at different developmental stages (Results). In addition to WGS, the elevated C>T mutation rate in CpG sites was also reported in high-fraction pSNMs identified from whole-exome sequencing data [11, 46].

Two of the reasons why the study of postzygotic mutations in healthy organs lags behind that of tumors include the lack of matched control samples in healthy individuals and the significantly lower abundance of postzygotic mutations. Our results showed that 38% of the validated pSNMs were shared by more than one organ, proving the importance of using a control-free pSNM-caller such as MosaicHunter. Furthermore, the high specificity of MosaicHunter compared to other callers enabled us to generate a candidate list that was specific enough to be validated. Single-cell sequencing has been demonstrated to be an alternative approach to study postzygotic mutations [14, 15]. However, compared to single-cell sequencing, bulk sequencing is able to not only provide the genomic location of the postzygotic mutations but also their allele fractions (Fig 1B), which are informative for assessing the proportion of cells that carry the mutation as well as reconstructing the lineage similarity across multiple samples within an individual (Fig 2E and 2F).

The list of pSNMs that had arisen locally during clonal expansion events in the liver and breast samples identified in our study deserve further discussion here. A cell clone with fitness advantage can predominantly proliferate faster and drive all private mutations that were originally carried by that clone to higher allele fractions, allowing them to be detected by bulk sequencing [26]. Because early embryonic pSNMs might affect only a fraction of cells in a certain organ, clonal expansion events could, in theory, make some pSNMs become undetectable from bulk sequencing if the carrier clone was out-competed. Indeed, we observed the breast #7 and breast #9 samples of BBL11121 had lost nine pSNMs that were detected in other breast samples and other organs of the same individual (Fig 2D). These results demonstrated the dynamics of allele fraction for pSNMs driven by clonal expansion events in healthy individuals. We further screened 1407 cancer-related genes from BBLD1005's liver samples using panel sequencing, and identified four more pSNMs with allele fraction around 1% (S4 Table). However, none of the clonal expansion pSNMs had been previously reported in cancer studies and more functional experiments might be required to examine their relationship with clonal expansion.

The current ~100X WGS bulk sequencing data in our study might not provide enough sensitivity to detect the whole spectrum of pSNMs, especially for those with allele fractions less than 1%. The genomic pattern we reported here were based on the analysis of eight organ types from five individuals. With reduced cost of NGS technology, we can expect a better-characterized spectrum of pSNMs in more and more organ samples and individuals in the future. A combination of deeper bulk sequencing and single-cell sequencing on the same organ sample could provide additional insights for pSNMs with lower allele fractions or even those present in only one or a few cells. This will enable a better characterization of postzygotic mutations in the human population and shed new light on distinguishing clinically-relevant postzygotic mutations from the genomic background.

Methods

Sample collection and preparation

Twenty-seven postmortem organ samples from five donors were obtained from BioServe Bio-technologies (Beltsville, MD, USA), with an approved protocol from the Institutional Review Board of Medical Informatics Multi-Media Systems, Inc. (5-2-07) and written informed consent obtained from all participants or their legal guardians (Table 1). The clinical histories of all five donors showed no diagnosis of cancer or other known overgrowth disorders. Each organ sample was dissected into nine pieces (roughly 0.5×0.5×0.5 cm each) perpendicular to its long axis and labeled from #1 to #9 (S7 Fig). The peripheral blood samples of two unrelated clinically unremarkable individuals of Asian descent (ACC1 and ACC4) were collected with written informed consent and approval by the Institutional Review Board at Peking University (IRB00001052-13025). Genomic DNA was extracted using an AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) after homogenization.

Whole-genome sequencing

Genomic DNA extracted from one piece (labeled as #9) of each of the 27 organ samples was used for WGS and subsequent validation. To reduce potential bias introduced by library preparation, three sequencing libraries were constructed independently for each sample using a KAPA LTP Library Preparation Kit for Illumina platforms (Kapa Biosystems, Wilmington, MA, USA). Size selection was performed for each library with a target insert size of 350–450 bp using a Pippin Prep system (Sage Science, Beverly, MA, USA). Libraries were purified using Agencourt AMPure XP beads (1.0× volume; Beckman Coulter, Brea, CA, USA) and underwent subsequent quality control using a 2100 Bioanalyzer (Santa Clara, CA, USA). Each library was sequenced in one lane on an Illumina Hiseq X Ten platform (Illumina, San Diego, CA, USA) using 150-bp paired-end reads.

Identification of postzygotic single-nucleotide mosaicisms

Raw sequencing reads were aligned to the GRCh37 human reference genome using the paired-end mode of BWA [51]. The aligned reads were processed using Picard and GATK [52] for the removal of duplicated and ambiguous reads (mismatch >4), indel realignment, and base-quality recalibration. The average depths of processed reads in each organ sample ranged from 83X to 113X (Table 1). CNVs and indels were called using CNVnator [53] and GATK [52], respectively, and all the involved regions as well as annotated repetitive regions were masked, because these regions were more vulnerable to false positives due to mis-alignment or abnormal copy numbers. To maximize the detection sensitivity of pSNMs, the single-sample and paired-sample modes of MosaicHunter [35] were applied to each organ sample with default parameters of genotyper and filters. For the paired-sample mode, the WGS data of the other organs obtained from the same donor served as paired controls. Candidates with at least 5% mutant allele fractions and 3 reads supporting the mutant allele were considered in our subsequent analyses. We randomly chose 16 candidates that are present in the latest versions of dbSNP [54] or the 1000 Genomes Project [55] to validate. As a result, all of them were genotyped as heterozygous rather than pSNMs, suggesting that they were more likely to be caused due to the large variation of allele fraction in WGS. Therefore, we only considered the candidates absent in both databases for thorough validation below.

Validation of postzygotic single-nucleotide mosaicisms

All the candidate pSNMs were analyzed using the standard workflow of PASM, which had been previously benchmarked by pyrosequencing and micro-droplet digital PCR [5, 56]. PCR primers for PASM were successfully designed for 241 out of the 251 candidate pSNMs, with their amplicon lengths ranging between 380 and 420 bps; for a few candidates located in highly homologous genomic regions, two rounds of nested PCR were carried out to achieve higher amplification specificity (S5 Table). For one-step amplicons, 35 cycles of PCR amplification was carried out using 2X Ex-Taq Premix (Takara Bio, Dalian, China). For nested amplicons, additional 15 cycles of first-round amplification and 25 cycles of second-round amplification was carried out to capture the target region specifically.

Amplicons from different organ samples and individuals were barcoded during library preparation and then pooled and sequenced using either the Ion Torrent PGM or Ion S5 XL sequencer (ThermoFisher, Guilford, CT, USA), following the manufacturer's protocols. Sequencing of the same PASM library on the PGM and S5 XL platforms showed a similar distribution of depth-of-coverage (S16 Fig). In each sample, we calculated the 95% credible intervals of minor allele fractions for each candidate site with at least 30X depth [5]. Candidates with 95% credible intervals between 0.5% and 40% were considered a mosaic genotype, whereas those sites with a 95% credible intervals' lower bound below 0.5% or an upper bound above 40% were considered a homozygous or heterozygous genotype, respectively. Five candidate sites without sufficient depth in the negative control samples were removed. As shown in S17 Fig, the inter-organ and intra-organ variation of the minor allele fractions for pSNMs could not be explained by the technical variance induced by DNA extraction or PASM.

The differences in pSNM profiles across multiple samples from one individual were assessed using the Euclidean distance (D) of the square root of the allele fraction for all the validated sites. The relative similarity (RS) between sample i and j was defined as

$$RS_{ij} = 1 - \frac{D_{ij} - \min(D)}{\max(D) - \min(D)}$$

Copy number estimation of postzygotic single-nucleotide mosaicisms

CNVkit [57] was applied to estimate copy number from the WGS data, and the 100 bp windows centered by each pSNM were considered. The individual-specific copy number was then normalized using the mean across all the five individuals. Three of the 167 PASM-validated pSNMs were found to demonstrate abnormal copy number (< 1.7 or > 2.3 for autosomes and females' X chromosome and < 0.9 or > 1.1 for males' X chromosome) in the corresponding carriers (S1 Table). The estimated copy numbers were 2.79, 2.80, and 2.59 for these three pSNMs (Q7, P35, and N26), respectively. Their mutant alleles were globally present in all the sequenced organs of the carrier, and the average allele fractions were close to 1/3 (26.2% to 30.9%). Rather than postzygotic mutations, these three sites were very likely to be explained by the germ-line events of copy number gain which made the allele fraction of involved heterozygous mutations deviate from 50%. Therefore, we excluded these three sites from all the following analyses.

Genomic annotation of replication timing, DNase sensitivity, and chromatin status

The genome-wide annotation of DNA replication timing was extracted from two independent studies [43, 58]. For Hansen *et al.*, we downloaded the wavelet-smoothed signal datasets of five

different cell-lines, including GM12878, K562, HeLa-S1, HepG2, and HUVEC [58]. For Koren *et al.*, the genome-wide profile was averaged from six lymphoblastoid cell-lines [43]. In both studies, a higher value represented an earlier DNA replication timing. For the permutation test, because our MosaicHunter pipeline only considered candidate pSNMs in non-repetitive regions, we compared the observed median of replication timing for each type of pSNMs against the null distribution estimated by using 1000 times of genome-wide random shuffling among non-repetitive regions.

The DNase-seq data of the embryonic stem cell-line H1-hESC was downloaded from the ENCODE project [59], and the DNase I sensitive zones were identified using the HotSpot algorithm [60]. For the permutation test, we randomly permutated the genomic positions of pSNMs among non-repetitive regions and assessed the median distance between each permutated pSNM and its closest DNase I sensitive zone. Each permutation was replicated 1000 times to estimate the distribution under the null hypothesis.

The annotation of chromatin states in HepG2, HMEC, and K562 cell-lines was downloaded from the UCSC Genome Browser [61] to represent cell types derived from the three primary germ layers in early embryo. Chromatin states were inferred from ChIP-seq data of ten epigenetic factors using a Hidden Markov Model [62]. The inferred chromatin states of active promoter, enhancer, and transcription were defined as "transcribed" chromatin status, whereas the other states involving repressed elements and heterochromatin were defined as "repressed" chromatin status.

The annotation of TADs for the H1-hESC cell-line was downloaded from the ENCODE project [59], which was generated based on the Hi-C data of Dixon *et al.* [63]. The expression profile of the same cell-line was also downloaded from the ENCODE project under GEO accession number GSM958733. A gene was defined as embryonically transcribed if its FPKM was larger than 20 in H1-hESC. We also used other FPKM thresholds to confirm the robustness of our finding.

Cancer panel sequencing of BBLD1005's liver samples

To screen for potential driver mutations related to the clonal expansion events in BBLD1005's liver samples, we captured 1407 cancer-related genes from the liver #2 and #8 samples of BBLD1005 using a Roche SeqCap panel (Pleasanton, CA, USA) designed by Genecast Biotechnology (Beijing, China). The captured libraries were sequenced by Illumina Novaseq6000 (Illumina, San Diego, CA, USA) using 150-bp paired-end reads, with an average depth of ~2000X. Sequencing reads were aligned to the GRCh37 human reference genome by BWA. The pSNMs were identified using MosaicHunter, and the 95% credible intervals of mutant allele fraction were estimated by the Bayesian model implemented in PASM.

Supporting information

S1 Fig. Distribution of per-site depth-of-coverage for PGM Amplicon Sequencing of Mosaicism (PASM). The average depth was greater than 4000X and more than 90% of the candidate sites were sequenced with at least 200X coverage in each sample.
(PDF)

S2 Fig. Mutant allele fractions and genotypes of candidate sites in multiple organs of BBL1100C. ACC1 and ACC4 are two unrelated individuals that served as negative controls. The mutant allele fractions were assessed by PGM Amplicon Sequencing of Mosaicism (PASM). Red, green, and blue colors denote heterozygous, mosaic, and reference-homozygous

genotypes, respectively.
(PDF)

S3 Fig. Mutant allele fractions and genotypes of candidate sites in multiple organs of BBL11121. ACC1 and ACC4 are two unrelated individuals that served as negative controls. The mutant allele fractions were assessed by PGM Amplicon Sequencing of Mosaicism (PASM). Red, green, and blue colors denote heterozygous, mosaic, and reference-homozygous genotypes, respectively.
(PDF)

S4 Fig. Mutant allele fractions and genotypes of candidate sites in multiple organs of BBLC1013. ACC1 and ACC4 are two unrelated individuals that served as negative controls. The mutant allele fractions were assessed by PGM Amplicon Sequencing of Mosaicism (PASM). Red, green, and blue colors denote heterozygous, mosaic, and reference-homozygous genotypes, respectively.
(PDF)

S5 Fig. Mutant allele fractions and genotypes of candidate sites in multiple organs of BBLD1005. ACC1 and ACC4 are two unrelated individuals that served as negative controls. The mutant allele fractions were assessed by PGM Amplicon Sequencing of Mosaicism (PASM). Red, green, and blue colors denote heterozygous, mosaic, and reference-homozygous genotypes, respectively.
(PDF)

S6 Fig. Mutant allele fractions and genotypes of candidate sites in multiple organs of BBLD1010. ACC1 and ACC4 are two unrelated individuals that served as negative controls. The mutant allele fractions were assessed by PGM Amplicon Sequencing of Mosaicism (PASM). Red, green, and blue colors denote heterozygous, mosaic, and reference-homozygous genotypes, respectively.
(PDF)

S7 Fig. Strategy for intra-organ multi-sampling. Targeted ultra-deep resequencing was performed on three additional liver samples (liver #2, #5, and #8) with varied physical distances to the original whole-genome sequenced liver sample (liver #9).
(PDF)

S8 Fig. Mutant allele fractions and genotypes across multiple liver samples of BBLD1005. The mutant allele fractions were assessed by PGM Amplicon Sequencing of Mosaicism (PASM). Red, green, and blue colors denote heterozygous, mosaic, and reference-homozygous genotypes, respectively.
(PDF)

S9 Fig. Mutant allele fractions and genotypes across multiple breast samples of BBL11121. The mutant allele fractions were assessed by PGM Amplicon Sequencing of Mosaicism (PASM). Red, green, and blue colors denote heterozygous, mosaic, and reference-homozygous genotypes, respectively.
(PDF)

S10 Fig. Distribution of allele fractions for clonal expansion pSNMs in liver and breast samples. Each sample showed a single peak for the mosaic allele fraction, suggesting that they originated from clonal expansion.
(PDF)

S11 Fig. Replication timing for pSNMs with varied allele fractions. The grey line denotes the genomic average. Embryonic pSNMs with a wide range of allele fractions contributed to the enrichment of early-replicating regions.
(PDF)

S12 Fig. Enriched embryonic pSNMs in the topologically associating domains (TADs) containing embryonically-transcribed genes. The X axis denotes different FPKM thresholds to define embryonically-transcribed genes, and the Y axis denotes the odds ratio of enrichment between embryonic and non-embryonic pSNMs. The odds ratios were robustly greater than one with varied FPKM thresholds.
(PDF)

S13 Fig. Chromatin status of tissue-shared and clone-specific pSNMs that were previously identified in Bae *et al*. The pSNMs shared across multiple brain regions or non-brain tissues showed significantly higher proportion of transcribed chromatin status than those specifically present in the clone of neuronal progenitor cells.
(PDF)

S14 Fig. Mutation spectrum for the embryonic pSNMs that were globally present in all the sequenced organs of the individual and those only present in some but not all the sequenced organs. Both types of embryonic pSNMs showed predominant C>T mutations and elevated mutation rate in CpG sites.
(PDF)

S15 Fig. Replication timing for the embryonic pSNMs that were globally present in all the sequenced organs of the individual and those only present in some but not all the sequenced organs. Both types of embryonic pSNMs demonstrated the similar enrichment for early-replicating regions.
(PDF)

S16 Fig. Similar distribution of depth-of-coverage between two platforms used for PGM Amplicon Sequencing of Mosaicism (PASM). The red and blue curves denote sequencing runs of the same library performed using an Ion Torrent PGM and an Ion S5 XL, respectively.
(PDF)

S17 Fig. Varied mosaic allele fractions across different samples of the same organ and different organs obtained from the same individual. Sample IDs starting with "b", "c", "l", "p", and "s" denote multiple samples obtained from brain, colon, liver, prostate, and skin of BBL1100C, respectively. Triple technical replicates of PGM Amplicon Sequencing of Mosaicism (PASM) were performed for the V18 pSNM of each sample. The inter-organ and intra-organ variations are dramatically larger than the variation observed between technical replicates.
(PDF)

S1 Table. Copy number estimation of candidate mosaic sites.
(DOC)

S2 Table. List of the 164 validated mosaic sites identified from five individuals.
(DOC)

S3 Table. Chromatin state for the embryonic pSNMs that were globally present in all the sequenced organs of the individual and those only present in some but not all the sequenced organs.
(DOC)

S4 Table. Mosaic sites identified by cancer panel sequencing of BBLD1005's liver samples.
(DOC)

S5 Table. Primers used for PGM Amplicon Sequencing of Mosaicism (PASM).
(DOC)

Author Contributions

Conceptualization: August Yue Huang, Liping Wei.

Data curation: August Yue Huang, Xianing Zheng.

Formal analysis: August Yue Huang.

Funding acquisition: Liping Wei.

Investigation: August Yue Huang.

Methodology: August Yue Huang, Xiaoxu Yang, Sheng Wang, Xianing Zheng.

Project administration: August Yue Huang, Qixi Wu, Liping Wei.

Resources: Sheng Wang, Qixi Wu.

Software: Adam Yongxin Ye.

Supervision: August Yue Huang, Liping Wei.

Validation: Xiaoxu Yang.

Visualization: August Yue Huang.

Writing – original draft: August Yue Huang, Xiaoxu Yang.

Writing – review & editing: August Yue Huang, Liping Wei.

References

1. Lupski JR. Genetics. Genome mosaicism—one human, multiple genomes. Science. 2013; 341 (6144):358–9. PMID: 23888031.

2. Poduri A, Evrony GD, Cai X, Walsh CA. Somatic mutation, genomic variation, and neurological disease. Science. 2013; 341(6141):1237758. https://doi.org/10.1126/science.1237758 PMID: 23828942.

3. Frank SA. Somatic mosaicism and disease. Curr Biol. 2014; 24(12):R577–81. https://doi.org/10.1016/j.cub.2014.05.021 PMID: 24937287.

4. Youssoufian H, Pyeritz RE. Mechanisms and consequences of somatic mosaicism in humans. Nat Rev Genet. 2002; 3(10):748–58. https://doi.org/10.1038/nrg906 PMID: 12360233.

5. Xu X, Yang X, Wu Q, Liu A, Yang X, Ye AY, et al. Amplicon Resequencing Identified Parental Mosaicism for Approximately 10% of "de novo" SCN1A Mutations in Children with Dravet Syndrome. Hum Mutat. 2015; 36(9):861–72. Epub 2015/06/23. https://doi.org/10.1002/humu.22819 PMID: 26096185.

6. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014; 505(7484):495–501. https://doi.org/10.1038/nature12912 PMID: 24390350.

7. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nat Rev Cancer. 2004; 4(3):177–83. https://doi.org/10.1038/nrc1299 PMID: 14993899.

8. Lindhurst MJ, Sapp JC, Teer JK, Johnston JJ, Finn EM, Peters K, et al. A mosaic activating mutation in AKT1 associated with the Proteus syndrome. N Engl J Med. 2011; 365(7):611–9. https://doi.org/10.1056/NEJMoa1104017 PMID: 21793738.

9. Couto JA, Huang AY, Konczyk DJ, Goss JA, Fishman SJ, Mulliken JB, et al. Somatic MAP2K1 Mutations Are Associated with Extracranial Arteriovenous Malformation. Am J Hum Genet. 2017; 100 (3):546–54. Epub 2017/02/14. https://doi.org/10.1016/j.ajhg.2017.01.018 PMID: 28190454.

10. Freed D, Pevsner J. The Contribution of Mosaic Variants to Autism Spectrum Disorder. PLoS Genet. 2016; 12(9):e1006245. https://doi.org/10.1371/journal.pgen.1006245 PMID: 27632392.

11. Dou Y, Yang X, Li Z, Wang S, Zhang Z, Ye AY, et al. Postzygotic single-nucleotide mosaicisms contribute to the etiology of autism spectrum disorder and autistic traits and the origin of mutations. Hum Mutat. 2017; 38(8):1002–13. Epub 2017/05/16. https://doi.org/10.1002/humu.23255 PMID: 28503910.

12. Holstege H, Pfeiffer W, Sie D, Hulsman M, Nicholas TJ, Lee CC, et al. Somatic mutations found in the healthy blood compartment of a 115-yr-old woman demonstrate oligoclonal hematopoiesis. Genome Res. 2014; 24(5):733–42. https://doi.org/10.1101/gr.162131.113 PMID: 24760347.

13. Huang AY, Xu X, Ye AY, Wu Q, Yan L, Zhao B, et al. Postzygotic single-nucleotide mosaicisms in whole-genome sequences of clinically unremarkable individuals. Cell Res. 2014; 24(11):1311–27. Epub 2014/10/15. https://doi.org/10.1038/cr.2014.131 PMID: 25312340.

14. Bae T, Tomasini L, Mariani J, Zhou B, Roychowdhury T, Franjic D, et al. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. Science. 2018; 359(6375):550–5. Epub 2017/12/09. https://doi.org/10.1126/science.aan8690 PMID: 29217587.

15. Lodato MA, Rodin RE, Bohrson CL, Coulter ME, Barton AR, Kwon M, et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. Science. 2018; 359(6375):555–9. Epub 2017/12/09. https://doi.org/10.1126/science.aao4426 PMID: 29217584.

16. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014; 371(26):2477–87. https://doi.org/10.1056/NEJMoa1409405 PMID: 25426838.

17. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014; 371(26):2488–98. https://doi.org/10.1056/NEJMoa1408617 PMID: 25426837.

18. Yadav VK, DeGregori J, De S. The landscape of somatic mutations in protein coding genes in apparently benign human tissues carries signatures of relaxed purifying selection. Nucleic Acids Res. 2016; 44(5):2075–84. https://doi.org/10.1093/nar/gkw086 PMID: 26883632.

19. Stratton MR. Exploring the genomes of cancer cells: progress and promise. Science. 2011; 331 (6024):1553–8. https://doi.org/10.1126/science.1204040 PMID: 21436442.

20. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet. 2014; 15(9):585–98. https://doi.org/10.1038/nrg3729 PMID: 24981601.

21. Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, et al. Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. Cell. 2016; 164 (3):538–49. https://doi.org/10.1016/j.cell.2015.12.050 PMID: 26806129.

22. Roberts SA, Gordenin DA. Hypermutation in human cancer genomes: footprints and mechanisms. Nat Rev Cancer. 2014; 14(12):786–800. https://doi.org/10.1038/nrc3816 PMID: 25568919.

23. Forsberg LA, Gisselsson D, Dumanski JP. Mosaicism in health and disease—clones picking up speed. Nat Rev Genet. 2017; 18(2):128–42. https://doi.org/10.1038/nrg.2016.145 PMID: 27941868.

24. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017; 168(4):613–28. Epub 2017/02/12. https://doi.org/10.1016/j.cell.2017.01.018 PMID: 28187284.

25. Aktipis CA, Boddy AM, Gatenby RA, Brown JS, Maley CC. Life history trade-offs in cancer evolution. Nat Rev Cancer. 2013; 13(12):883–92. https://doi.org/10.1038/nrc3606 PMID: 24213474.

26. Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015; 348 (6237):880–6. PMID: 25999502.

27. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. Nat Rev Genet. 2013; 14(5):307–20. https://doi.org/10.1038/nrg3424 PMID: 23594909.

28. Watson IR, Takahashi K, Futreal PA, Chin L. Emerging patterns of somatic mutations in cancer. Nat Rev Genet. 2013; 14(10):703–18. https://doi.org/10.1038/nrg3539 PMID: 24022702.

29. Wang J, Fan HC, Behr B, Quake SR. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. Cell. 2012; 150(2):402–12. https://doi.org/10.1016/j.cell.2012.06.030 PMID: 22817899.

30. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, et al. Tissue-specific mutation accumulation in human adult stem cells during life. Nature. 2016; 538(7624):260–4. https://doi.org/10.1038/nature19768 PMID: 27698416.

31. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. Science. 2015; 350(6256):94–8. Epub 2015/10/03. https://doi.org/10.1126/science.aab1785 PMID: 26430121.

32. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. Nat Rev Genet. 2016; 17(3):175–88. Epub 2016/01/26. https://doi.org/10.1038/nrg.2015.16 PMID: 26806412.

33. Shapiro E, Biezuner T, Linnarsson S. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat Rev Genet. 2013; 14(9):618–30. https://doi.org/10.1038/nrg3542 PMID: 23897237.

34. Grun D, van Oudenaarden A. Design and Analysis of Single-Cell Sequencing Experiments. Cell. 2015; 163(4):799–810. https://doi.org/10.1016/j.cell.2015.10.039 PMID: 26544934.

35. Huang AY, Zhang Z, Ye AY, Dou Y, Yan L, Yang X, et al. MosaicHunter: accurate detection of postzygotic single-nucleotide mosaicism through next-generation sequencing of unpaired, trio, and paired samples. Nucleic Acids Res. 2017; 45(10):e76. Epub 2017/01/31. https://doi.org/10.1093/nar/gkx024 PMID: 28132024.

36. Lynch M. Evolution of the mutation rate. Trends Genet. 2010; 26(8):345–52. https://doi.org/10.1016/j.tig.2010.05.003 PMID: 20594608.

37. Campbell IM, Yuan B, Robberecht C, Pfundt R, Szafranski P, McEntagart ME, et al. Parental somatic mosaicism is underrecognized and influences recurrence risk of genomic disorders. Am J Hum Genet. 2014; 95(2):173–82. https://doi.org/10.1016/j.ajhg.2014.07.003 PMID: 25087610.

38. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500(7463):415–21. https://doi.org/10.1038/nature12477 PMID: 23945592.

39. Ju YS, Martincorena I, Gerstung M, Petljak M, Alexandrov LB, Rahbari R, et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. Nature. 2017; 543(7647):714–8. Epub 2017/03/23. https://doi.org/10.1038/nature21703 PMID: 28329761.

40. Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, Tamuri AU, et al. Genome sequencing of normal cells reveals developmental lineages and mutational processes. Nature. 2014; 513(7518):422–5. https://doi.org/10.1038/nature13448 PMID: 25043003.

41. Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. Nat Genet. 2009; 41(4):393–5. Epub 2009/03/17. https://doi.org/10.1038/ng.363 PMID: 19287383.

42. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013; 499(7457):214–8. https://doi.org/10.1038/nature12213 PMID: 23770567.

43. Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. Am J Hum Genet. 2012; 91(6):1033–40. https://doi.org/10.1016/j.ajhg.2012.10.018 PMID: 23176822.

44. Makova KD, Hardison RC. The effects of chromatin organization on variation in mutation rates in the genome. Nat Rev Genet. 2015; 16(4):213–23. Epub 2015/03/04. https://doi.org/10.1038/nrg3890 PMID: 25732611.

45. Servick K. Clinical testing. Harmful mutations can fly under the radar. Science. 2014; 345(6203):1438–9. PMID: 25237080.

46. Lim ET, Uddin M, De Rubeis S, Chan Y, Kamumbu AS, Zhang X, et al. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. Nat Neurosci. 2017; 20(9):1217–24. Epub 2017/07/18. https://doi.org/10.1038/nn.4598 PMID: 28714951.

47. Supek F, Lehner B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. Nature. 2015; 521(7550):81–4. https://doi.org/10.1038/nature14173 PMID: 25707793.

48. Abyzov A, Tomasini L, Zhou B, Vasmatzis N, Coppola G, Amenduni M, et al. One thousand somatic SNVs per skin fibroblast cell set baseline of mosaic mutational load with patterns that suggest proliferative origin. Genome Res. 2017; 27(4):512–23. Epub 2017/02/27. https://doi.org/10.1101/gr.215517.116 PMID: 28235832.

49. Saini N, Roberts SA, Klimczak LJ, Chan K, Grimm SA, Dai S, et al. The Impact of Environmental and Endogenous Damage on Somatic Mutation Load in Human Skin Fibroblasts. PLoS Genet. 2016; 12(10):e1006385. Epub 2016/10/28. https://doi.org/10.1371/journal.pgen.1006385 PMID: 27788131.

50. Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. Differences between germline and somatic mutation rates in humans and mice. Nat Commun. 2017; 8:15183. Epub 2017/05/10. https://doi.org/10.1038/ncomms15183 PMID: 28485371.

51. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168.

52. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43(5):491–8. https://doi.org/10.1038/ng.806 PMID: 21478889.

53. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011; 21(6):974–84. https://doi.org/10.1101/gr.114876.110 PMID: 21324876.

54. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001; 29(1):308–11. PMID: 11125122.

55. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491(7422):56–65. https://doi.org/10.1038/nature11632 PMID: 23128226.

56. Yang X, Liu A, Xu X, Yang X, Zeng Q, Ye AY, et al. Genomic mosaicism in paternal sperm and multiple parental tissues in a Dravet syndrome cohort. Sci Rep. 2017; 7(1):15677. Epub 2017/11/17. https://doi.org/10.1038/s41598-017-15814-7 PMID: 29142202.

57. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Comput Biol. 2016; 12(4):e1004873. Epub 2016/04/23. https://doi.org/10.1371/journal.pcbi.1004873 PMID: 27100738.

58. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A. 2010; 107(1):139–44. https://doi.org/10.1073/pnas.0912402107 PMID: 19966280.

59. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489 (7414):57–74. https://doi.org/10.1038/nature11247 PMID: 22955616.

60. Sabo PJ, Hawrylycz M, Wallace JC, Humbert R, Yu M, Shafer A, et al. Discovery of functional noncoding elements by digital analysis of chromatin structure. Proc Natl Acad Sci U S A. 2004; 101(48):16837–42. Epub 2004/11/20. https://doi.org/10.1073/pnas.0407387101 PMID: 15550541.

61. Dreszer TR, Karolchik D, Zweig AS, Hinrichs AS, Raney BJ, Kuhn RM, et al. The UCSC Genome Browser database: extensions and updates 2011. Nucleic Acids Res. 2012; 40(Database issue):D918–23. https://doi.org/10.1093/nar/gkr1055 PMID: 22086951.

62. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473(7345):43–9. Epub 2011/03/29. https://doi.org/10.1038/nature09906 PMID: 21441907.

63. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485(7398):376–80. Epub 2012/04/13. https://doi.org/10.1038/nature11082 PMID: 22495300.