Somatic mutations in single human cardiomyocytes demonstrate accelerated age-related DNA damage and cell fusion

Christopher Walsh (Christopher.Walsh@childrens.harvard.edu)
Boston Children's Hospital https://orcid.org/0000-0002-0156-2238
Sangita Choudhury
Boston Childrens hospital
August Huang
Boston Childrens hospital
Junho Junho Kim
Boston Childrens hospital
Katherine Morillo
Boston Childrens Hospital
Eduardo Maury
Boston Childrens hospital
Zinan Zinan Zhou
Boston Childrens hospital
Connor Kenny
Boston Childrens hospital
Eunjung Lee
Boston Childrens hospital
Ming Chen
Boston Childrens hospital

Article

Keywords: Cardiac Mutations, Polyploidy, Mutational Signatures, Lineage Tree, Genomic Aging

Posted Date: October 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-84503/v1

Somatic mutations in single human cardiomyocytes demonstrate accelerated age-related DNA damage and cell fusion

Sangita Choudhury[1*], August Yue Huang[1*], Junho Kim[1], Katherine Morillo[1], Eduardo A. Maury[1, 2], Zinan Zhou[1], Connor Kenny[1], Eunjung Alice Lee[1]†, Ming Hui Chen[1, 3]†, Christopher Walsh[1, 4]†

1. Division of Genetics and Genomics, Manton Center for Orphan Disease, Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA; Departments of Neurology and Pediatrics, Harvard Medical School, Boston, MA, USA
2. Bioinformatics & Integrative Genomics Program and Harvard/MIT MD-PHD Program, Harvard Medical School, Boston, MA
3. Department of Cardiology, Boston Children's Hospital, Boston, MA, USA; Department of Cardiology, Harvard Medical School, Boston, MA, USA
4. Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA

* These authors contributed equally to this work

† Lead contact and corresponding author:

Christopher A Walsh

3 Blackfan St, CLS-15064

Boston, MA 02115

Phone: (617) 919-2923 | Fax: (617) 919-2010

christopher.walsh@childrens.harvard.edu.

Ming Hui Chen, MD, MMSc, FACC

300 Longwood Ave

Boston, MA 02115

Phone (617) 355-0652

Minghui.chen@cardio.chboston.org

Eunjung Alice Lee

3 Blackfan St, CLS-15020

Boston, MA 02115

Phone: (617) 919-1589 | Fax: (617) 919-2923

ealice.lee@childrens.harvard.edu

Introductory Paragraph:

The accumulation of somatic DNA mutations is a hallmark of aging in many dividing cells and contributes to carcinogenesis[1-7], but non-cycling cells have been thought to be protected from mutation. Here we survey the landscape of somatic single-nucleotide variants (sSNVs) in heart muscle cells (cardiomyocytes) which normally do not proliferate[8-11] but often become polyploid with age[12-14]. Using single-cell whole-genome sequencing[1,15,16] we analyzed sSNVs from 48 single cardiomyocytes from 10 healthy individuals (ages 0.4 - 82 yrs.). Cardiomyocyte sSNVs increased strikingly with age, at rates faster than reported in many dividing cells, or in nondividing neurons[1 5,17,18]. Analysis of nucleotide substitution patterns revealed age-related "clock-like" mutational signatures resembling those previously described[15 19,20]. However, cardiomyocytes showed distinct mutational signatures that are rare or absent in other cells, implicating failed nucleotide excision repair of oxidative damage and defective mismatch repair (MMR) during aging. A lineage tree of cardiomyocytes, constructed using clonal sSNVs, revealed that some tetraploid (10%) and most cardiomyocytes with higher ploidy (>60%) derive from distinct clonal origins, implicating cell fusion as a mechanism contributing to many polyploid cardiomyocytes. Since age-accumulated sSNVs create dozens of damaging exonic mutations, cell fusion to form multiploid cardiomyocytes may represent an evolutionary mechanism of cellular genetic compensation that minimizes complete knockout of essential genes during aging. The rates and patterns of accumulation of cardiac mutations provide a paradigm to understand the influence of genomic aging on age-related heart disease.

Introduction:

Advanced age is the most important risk factor for heart disease and heart failure[21,22,23,24]. Most heart failure occurs in individuals over age 65[22,25,26] and yet we have an incomplete understanding of how aging promotes heart failure. The accumulation of somatic DNA mutations recently has been demonstrated to be a hallmark of aging in many human cell types[3,5,15,27]. Recent evidence suggests that environmental factors causing somatic mutations may play a role in cancer as well as in other common diseases[4,6,15], including heart failure[28-30]. With aging, cardiac function is impaired at the organ level as well as the cardiomyocyte level[11,31]. In the adult heart, cardiomyocyte cell size is stable and proliferation is either very rare or absent[9,10,14,32]. Nevertheless, cardiomyocytes show changes in activation, contraction and relaxation, hypertrophy, and in some case, cell loss[33]. The underlying cause of these many aging phenotypes is likely molecular in nature but its mechanism is not well understood.

An unusual feature of cardiomyocytes, compared with other tissues, is their tendency to become polyploid with age, accumulating more than one nucleus, as well as more than one set of diploid chromosomes[13,14]. Polyploidy in nature is considered to be beneficial, since gene redundancy shields polyploid cells from the deleterious effect of certain mutations or environmental stresses[34-36]. Endoreplication and cell fusion are two drivers for the generation of polyploid cells[37,38]. Although endoreplication has been considered to be the main source of polyploid cardiomyocyte generation, recent evidence indicates that cardiomyocyte fusion may play potential roles in zebrafish[39] and murine[40,41] cardiac development, though its role in aging is unknown. Cell lineage studies have been performed on experimental animals by prospectively labeling dividing cells with, for example, thymidine analogs and cre technology[42,43], but the use

of this strategy is quite challenging in humans. This prompted us to evaluate the mechanism of polyploidization and explore the landscape of sSNVs in aging human heart cardiomyocytes.

Results:

Cardiomyocyte polyploidization starts early in life and gradually increases with age

To understand the temporal dynamics of polyploidization in human cardiomyocytes, we systematically evaluated the ploidy of cardiomyocyte nuclei from more than 50 human heart tissue samples from individuals ranging from 44 days to 81 years of age. Previous studies have shown that human cardiomyocyte nuclei are mostly diploid at birth and start to become polyploid mainly in the second decade of life[8]. We purified cardiac nuclei from fresh-frozen, unfixed, human myocardium by density sedimentation[44]. Diploid cardiomyocyte nuclei were identified with cell-specific markers, cardiac troponin T. Polyploid cardiomyocyte nuclei were identified with the nuclear stain, 4′, 6-diamidino-2-phenylindole (DAPI) as well as cell-specific markers (Figure 1A, upper). Our systemic analysis revealed that ~5% of cardiomyocyte nuclei were tetraploid in the 44 day-old heart sample. By 20 years of age, ~35% of cardiac nuclei are tetraploid, and ~5% of the cardiomyocytes were hexaploid or even higher in ploidy (hexaploid+). The proportion of tetraploid and hexaploid+ cardiomyocytes significantly increased with age ($p = 1.5\times10^{-15}$ and 3.4×10^{-3}, respectively, linear regression, Figure 1B). Our data indicate that polyploidization does not start late in life, but polyploidization in cardiomyocytes is evident in neonatal heart and increases with age.

Cellular fusion is a significant driver of cardiomyocyte polyploidization

Numerous non-inherited somatic mutations record the unique history of each somatic cell originating from a zygote. The clonal structure of these somatic mutations enables us to reconstruct the developmental lineage of somatic cells[1]. In order to study the mechanism of polyploidization in human cardiomyocyte, we isolated a total of 482 single cardiomyocyte nuclei (182 diploid, 173 tetraploid, and 127 hexaploid+) from the postmortem left ventricle heart tissue of a 17-year-old donor using fluorescence-activated nuclear sorting (FANS). We then amplified the DNA from each nucleus by multiple displacement amplification (MDA)[16]. Ultra-deep targeted sequencing for 482 MDA-amplified nuclei libraries were performed at an average depth of >7000X (Supplementary Figure S1), using a panel of 253 validated clonal sSNVs from the joint analysis of single-cell and bulk whole-genome sequencing (WGS) data from the same donor[1,45] (Figure 1A, middle).

As a result, 75.7% (365/482) of the sequenced cardiomyocytes could be placed into one of the 9 distinct nested clades defined by the presence and absence of one or more independent mutations (Figure 1C-F, Supplementary Table S1) with uncladed cells likely reflecting incomplete genome amplification. Only 1/148 diploid cardiomyocyte carried sSNVs from multiple clades (Figure 1D), presumably reflecting sorting of two nuclei into the same well (doublets), showing that this rarely occurs during FANS. In contrast, 5.8% (8/138) of cladded tetraploid cardiomyocytes (Figure 1E, G) and 19.3% (16/83) of cladded hexaploid+ cardiomyocytes (Figure 1F, G) contained sSNVs from multiple clades. We observed three- or four-cladded cells only in the hexaploid+ cardiomyocyte population (Figure 1F, G). Since multi-clade cells can only be explained by cell fusion rather than endoreplication, our results suggest that at least ~10% of tetraploid cardiomyocytes are generated from the fusion mechanism, and this proportion increases up to ~63% in cardiomyocytes with higher ploidy, after adjusting for

the sensitivity of detecting sSNVs from different clades (see Methods, Supplementary Table S2). We note that even this estimate likely under-states the frequency of cell fusion, due to its limitation in detecting fusing cells that originate from relatively closely related clones.

Nonclonal somatic mutations increase with age in cardiomyocytes

We evaluated the genome-burden of sSNVs using single-cell WGS of 40 tetraploid and 8 diploid single cardiomyocyte nuclei from postmortem hearts of young (<4 yrs.), middle-age (30-63 yrs.), and aged individuals (>75 yrs.) (Figure 1A, bottom). Cardiomyocyte nuclei DNA was amplified using MDA[16] followed by quality control steps (see Methods) and high-coverage (>50X) WGS[15] for selected, well-amplified cells (Supplementary Table S3). We identified single-cell sSNVs and estimated their genome-wide burden for each cardiomyocyte (Supplementary Table S4) using a modified version of the LiRA[46] algorithm that takes cell-specific dropout rates and sequencing depth distributions into consideration to account for the tetraploidy effect on detection sensitivity and power calculation (see Methods). sSNVs originating from likely MDA amplification artifacts[47], which show a distinctive nucleotide substitution pattern, were also excluded.

Tetraploid cardiomyocytes from hearts of aged individuals showed an average of ~11,500 sSNVs per tetraploid genome, compared to ~2,600 sSNVs per tetraploid genome from young individuals, with an increasing rate of ~118 ± 22 [s.e.] sSNVs per year ($p = 2.0\times10^{-4}$, linear regression; Figure 2A). The age-dependent increase of sSNVs remained significant even after adjusting for the unevenness of MDA amplification (Supplementary Figure S2). The SNVs were distributed broadly across the genome (Figure 2B). Note that at each age group there are significant intra-individual and inter-individual variations, particularly in the aged group, where a

few outlier nuclei showed very high numbers (>15,000 per cell) of sSNVs. We further confirmed the age-dependent trend in diploid cardiomyocytes, in which the aged heart showed an average of ~3,900 sSNVs per diploid cardiomyocyte genome compared to ~600 sSNVs in the young heart ($p = 0.013$, two-tailed T-test; Figure 2C). Our analysis revealed increasing somatic mutational burden in human heart muscle cells with age, regardless of cardiomyocyte nuclear ploidy.

Further we compared the sSNV accumulation rate in cardiomyocytes with that in neurons[1], another non-dividing human cell type, and found that aged cardiomyocytes accumulated sSNVs much faster than neurons (~118 ± 22 [s.e.] sSNVs per year in tetraploid cardiomyocyte *vs* ~23 ± 7 [s.e.] sSNVs per year in neurons), after correcting for the unevenness of whole-genome amplification ($p < 6.2\times10^{-10}$ for tetraploid cardiomyocytes *vs* neurons and $p < 2.9\times10^{-3}$ for diploid cardiomyocytes vs neurons, linear regression; Figure 2D). Tetraploid cardiomyocytes showed a ~2.5-fold higher rate of sSNV increase per-diploid genome than neurons, after accounting for the doubled genomic size in tetraploid cardiomyocytes as compared with neurons. Our results suggest that cardiomyocytes and neurons have differing rates of mutational accumulation during aging.

Signature analysis identifies distinct mutational processes in aging cardiomyocytes

Patterns and types of DNA substitution mutations can be deconvoluted into "signatures" that provide important biological information about their potential causes[48]. Our mutational analysis from 40 tetraploid cardiac nuclei detected 8,197 base substitutions (Supplementary Table S5). First, we compared the base substitution pattern of age-accumulated sSNVs by subtracting the sSNV profiles of young cardiomyocytes from aged cardiomyocytes, and observed

that C>T and T>C mutations accumulated predominantly during the aging process (Figure 3A). To decipher mutational processes in heart aging, we performed non-negative matrix factorization[19] using the sSNV profiles of cardiomyocytes as described previously[1,15]. Cardiac sSNVs were best fit into four distinct molecular signatures referred to as Signature A, B, C, and D in this study (Supplementary Figure S3). Signature B consisted mainly of C>T mutations, depleted at CpG dinucleotides, and matched closely with mutational signatures that were previously ascribed to artifacts created by MDA amplification[47], and so this signature was removed and not considered further (Supplementary Figure S4). The remaining signatures (Figure 3B) were similar to known signatures in cancer, neurons, and normal replicating cells suggesting that they may reflect important biological sources of mutations and distinct from any known artifact signatures.

Signature A mainly comprised C>T and T>C mutations and its contribution in cardiomyocytes increased with age. Signature A resembles age-related, “clock-like" signatures previously observed in many cancers[19,49], normal cycling cells[20], and single neurons[15], proposed to reflect faulty repair of deamination of methylated cytosines to thymine that frequently occurs in the CpG context[50], and aging-associated (Figure 3A). Our data show a similar clock-like signature in post mitotic cardiomyocytes that is independent of *cell division*. Hierarchical clustering of Signature A showed that Signature A closely resembled COSMIC signature SBS5, known as an aging signature.

Signature C was distinct from Signature A due to the prominent enrichment of C>A mutations (Figure 3B). C>A mutations often reflect increased 8-oxoguanine levels, created by oxidative DNA damage[51,52], one of the most common threats to genome stability[53]. The heart is one of the most highly metabolic organs, with large oxidative demand resulting in reactive

oxygen species (ROS) production[54]. During myocardial remodeling (aging or myocardial dysfunction), ROS production might be deregulated, producing 8-oxoguanine (8-hydroxyguanine, 8-oxo-Gua, or OH8Gua), one of the most common DNA lesions[55], and causing mismatched pairing with adenine resulting in G to T and C to A substitutions in the genome. Oxygen radicals also react with 5-methylcytosine to result in C>T transitions. Signature C mutations could arise due to the defective nucleotide excision repair of oxidative lesions.

A third signature, Signature D was also enriched for C>T and T>C mutations but distinct from Signature A in its tri-nucleotide context (Figure 3B). Signature D closely resembled COSMIC Signature SBS44 (Figure 3C, Supplementary Figure S5), which is associated with defective DNA mismatch repair (MMR) machinery and increases in tumors associated with loss of mismatch repair genes[56,57]. Moreover, Signature D also showed higher similarity to six different defective MMR signatures SBS6, 14, 15, 20, 21, 26 (Supplementary Figure S6). MMR is regulated by a small set of MMR-specific proteins in all cells [58,59]. Mutagenesis in the absence of one of the core MMR factors is shaped by the sequence spectrum of the unrepaired mismatches, which themselves are the product of the insertional specificity and proofreading activity of DNA polymerases. Signature D in the heart showed a striking similarity to the mutational signature of MSH6 defective human HAP1 cells, or the DLD-1 human colorectal cancer cell line, dominated by C > T and T > C mutations[60] in a range of contexts, plus small contribution of C>A, C>G, T>A, and T>G mutations. Signature D hence likely reflects a defect in the repair of damage that involves almost all nucleotides.

Comparative signature analysis between cardiomyocytes and neurons—another non-dividing cell—shows shared and distinct mutational signatures (Figure 3D, E). sSNVs from cardiomyocytes and neurons share an indistinguishable clock-like aging Signature A with a

similar rate of age-related increase in the distinct cell types, and also a similar Signature C enriched in C>A substitutions, that also modestly increases with age (Supplementary Figure S7). On the other hand, the dramatic increase in the contribution of Signature D with age was observed only in cardiomyocytes, not in neurons (Figure 3D). sSNVs that accumulated in aging cardiomyocytes showed broader substitution types with enrichment in untranscribed strands, such as T>C and T>G mutations than those in neurons, whereas age-accumulated sSNVs in neurons showed the enrichment of C>T and T>C mutations in transcribed strands (Figure 3E). This putative MMR signature, which is relatively specific to aging cardiomyocytes, represents a distinct mutational process in the heart.

Potential source and mechanisms of mutation formation in the aging heart

To understand how mutations can be formed and accumulate during aging in the absence of cell cycling, it is important to recognize that both MMR and NER involve steps of exonuclease removal of a segment of DNA, followed by replication of the remaining strand to reconstitute double-stranded DNA. Hence, even nondividing cells are still undergoing DNA replication of small segments of the genome, so that errors of replication, or two nearby damaged based on opposite strands[61,62], could result in fixed, double stranded mutations. Our analysis of mutational signatures (Figure 3B, C) suggested a model where oxidative stress leads to an increased burden of damaged bases, which might overwhelm the MMR machinery of aged cardiomyocytes. Using GTEx expression data[63], we observed a significant decrease in gene expression for the core components of the MMR complex (*MLH1, MLH3, MSH2, MSH3, MSH6, PMS1* and *PMS2*) with aging in heart samples ($p = 0.01$, linear regression; Figure 5A), and

overall lower expression levels of these genes in heart versus brain samples ($p = 2.5\times10^{-8}$, two-tailed paired Wilcoxon test; Figure 4A).

We directly assessed potential oxidative damage in the heart using an ELISA (see Methods). We found that the level of 8-hydroxy-2'-deoxyguanosine in aged hearts was more than twice as high as in younger hearts ($p = 0.008$, two-tailed Wilcoxon test; Figure 4B). These data suggest that aging results in increased generation, decreased repair, or both, of oxidative DNA lesions. It is known that failure of DNA MMR is associated with a strikingly elevated rate of base substitution mutagenesis, as a consequence, tumors with MMR deficiency are amongst those with the highest number of somatic mutations[64]. The MSH2-MSH6 complex is primarily responsible for repairing single base-base mismatches. Our findings of increased sSNV count in aged heart suggest that MMR is not efficient at correcting mismatched nucleotides in aged cardiomyocytes, contributing to the large increase in sSNVs. Taken together, these findings indicate that increased oxidative stress leads to elevated 8-hydroxy-2-deoxyguanosine, which may overwhelm DNA repair systems resulting in increased DNA mutational burden.

Functional impact of sSNVs in the aging process

To further investigate whether the occurrence of sSNVs is associated with defective gene transcription and function, we stratified and compared cardiac and neuronal sSNVs by using the expression profiles of corresponding tissues in GTEx[63]. Our results suggest that Signature A sSNVs are statistically enriched in highly expressed genes at a similar level in both aged cardiomyocytes and aged neurons ($p < 0.01$, linear regression; Figure 4C, upper), suggesting Signature A as a common transcription-associated signature[5,49]. In contrast, Signatures C and D showed higher contributions in aged cardiomyocytes than in aged neurons, without a strong

association with gene transcription level ($p < 0.1$, linear regression; Figure 4C, middle and lower), indicating that Signatures C and D might result from mechanisms different from Signature A and specific to cardiomyocytes. The age-dependent accumulation of sSNVs in genes highly expressed in cardiomyocytes suggests that the very genes essential for the function of cardiomyocytes are those most likely to be damaged during one's lifetime. Gene Ontology (GO) analysis found that genes involved in mismatch repair, mitochondria organization, and PI3Kinase pathways showed sSNV enrichment (FDR-adjusted $p < 0.05$, permutation test; Figure 4D). We also identified 25 nonsynonymous and 1 stop-gain sSNVs in aged cardiomyocytes (Supplementary Table S6).

Although many heterozygous mutations would likely compromise cardiomyocyte function, it is expected that deleterious gene "knockout" (KO) mutations in genes would be especially damaging if all the alleles are affected, and that there may be a threshold for the accumulation of such KOs above which cardiomyocyte function would deteriorate. Whereas either endoreplication or fusion of genetically unrelated cells may potentially represent adaptive mechanisms to guard against these KO mutations, we compared these two by estimating the accumulation of gene KOs in cardiomyocytes using a probability model where at least two coincident deleterious sSNV events in a diploid cell, or four deleterious sSNVs within one gene in tetraploid cells, cause loss of function. Tetraploid cardiomyocytes generated through endoreplication had an average probability of 1% of getting one or more genes completely knocked out by age 60, with this probability increasing to 6% by age 80, implying that a substantial fraction of cardiomyocytes would carry damaging mutations (Figure 4E). In contrast, cardiomyocytes generated by fusion showed a significantly lower probability of gene KO ($p = 0.002$, two-tailed paired Wilcoxon test; Figure 4E), with less than 0.5% of cells with KO genes at

age of 80 (Figure 4E). These data strongly support that fusion is more effective mechanism to avert the loss of gene function caused by age-related mutations (Figure 4F). We note also that our WGS-based method does not identify the source of the fusing cells, thus we cannot distinguish fusion of cells from distantly or closely-related cardiomyocytes.

Our results showed that each human cardiomyocyte has a profoundly distinctive genome, harboring as many as 4000 to 30000 somatic SNVs. These estimates are likely to improve in accuracy with better amplification methods as well as a larger sample size. We also note that our samples are biased towards cells that amplified well and equally, *i.e.*, cells more likely to have intact genomes and so may underestimate mutation rates. Cardiomyocyte SNVs display signatures of mutagenic processes, including increased oxidative stress, defective mismatch repair and a preponderance of C>T deamination especially at the CpG context (Figure 4F). Our analysis of human cardiomyocytes lays a foundation for better understanding the genomic landscape and mechanisms driving mutation accumulation in aging cardiomyocytes that may help develop new treatments to reduce age-related cardiomyocyte dysfunction.

Acknowledgment:

We thank R. Sean Hill, Jennifer N. Partlow for assistance, Boston Children's Hospital Flow Cytometry Core, IDDRC Molecular Genetics Core Facility, and the Research Computing group at Harvard Medical School for assistance. Human tissue was obtained from the NIH NeuroBioBank at the University of Maryland, and we thank the donors and their families for their invaluable donations for the advancement of science. American Heart Association-SDG (S.C), the Manton Center for Orphan Disease Research (C.A.W., E.A.L), supported this work. C.A.W. is supported by the Allen Discovery Center for Human Brain Evolution, funded by the Paul G. Allen Frontiers Program, and is an Investigator of the Howard Hughes Medical Institute. This work is supported in part by the NINDS by a supplement to R01NS032457. E.A.L was supported by NIA K01 AG051791, Suh Kyungbae Foundation, the Paul G. Allen Family Foundation, and the Charles H. Hood Foundation. E.A.M is supported by (T32GM007753) National Institute of General Medical Sciences to the Harvard/MIT MD-PhD program, and T15LM007092 to the Biomedical Informatics and Data Science Research Training Program.

Author contributions:

S.C, M.H.C, A.Y.H, E.A.L, and C.A.W conceived and designed the study. C.A.W, E.A.L, and M.H.C supervised the study. S.C, A.Y.H, E.A.L, M.H.C, and C.A.W wrote the manuscript. S.C, K.M, C.K, and Z.Z isolated single nuclei and performed whole-genome amplification as well as single-cell sequencing. A.Y.H performed sSNV calling from single-cell sequencing data and performed bioinformatics analyses. J.K performed QC and read alignment of single-cell WGS data. S.C and A.Y.H. designed and A.Y.H. analyzed clonal somatic mutations for panel

sequencing. E.A.M and A.Y.H analyzed GTEx expression data and performed knock-out cell modeling. All authors reviewed and edited the manuscript.

Competing interest declaration:

The authors declare that they have no competing interests.

Reference:

1 Lodato, M. A. *et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94-98, doi:10.1126/science.aab1785 (2015).

2 Franco, I. *et al.* Whole genome DNA sequencing provides an atlas of somatic mutagenesis in healthy human cells and identifies a tumor-prone cell type. *Genome Biol* **20**, 285, doi:10.1186/s13059-019-1892-z (2019).

3 Franco, I. *et al.* Somatic mutagenesis in satellite cells associates with human skeletal muscle aging. *Nat Commun* **9**, 800, doi:10.1038/s41467-018-03244-6 (2018).

4 Martincorena, I. *et al.* Somatic mutant clones colonize the human esophagus with age. *Science* **362**, 911-917, doi:10.1126/science.aau3879 (2018).

5 Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538**, 260-264, doi:10.1038/nature19768 (2016).

6 Lee-Six, H. *et al.* The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* **574**, 532-537, doi:10.1038/s41586-019-1672-7 (2019).

7 McConnell, M. J. *et al.* Intersection of diverse neuronal genomes and neuropsychiatric disease: The Brain Somatic Mosaicism Network. *Science* **356**, doi:10.1126/science.aal1641 (2017).

8 Lazar, E., Sadek, H. A. & Bergmann, O. Cardiomyocyte renewal in the human heart: insights from the fall-out. *Eur Heart J* **38**, 2333-2342, doi:10.1093/eurheartj/ehx343 (2017).

9 Bergmann, O. *et al.* Evidence for cardiomyocyte renewal in humans. *Science* **324**, 98-102, doi:10.1126/science.1164680 (2009).

10 Bergmann, O. *et al.* Dynamics of Cell Generation and Turnover in the Human Heart. *Cell* **161**, 1566-1575, doi:10.1016/j.cell.2015.05.026 (2015).

11 Ahuja, P., Sdek, P. & MacLellan, W. R. Cardiac myocyte cell cycle control in development, disease, and regeneration. *Physiol Rev* **87**, 521-544, doi:10.1152/physrev.00032.2006 (2007).

12 Derks, W. & Bergmann, O. Polyploidy in Cardiomyocytes: Roadblock to Heart Regeneration? *Circ Res* **126**, 552-565, doi:10.1161/CIRCRESAHA.119.315408 (2020).

13 Han, L. *et al.* Lamin B2 Levels Regulate Polyploidization of Cardiomyocyte Nuclei and Myocardial Regeneration. *Dev Cell* **53**, 42-59 e11, doi:10.1016/j.devcel.2020.01.030 (2020).

14 Mollova, M. *et al.* Cardiomyocyte proliferation contributes to heart growth in young humans. *Proc Natl Acad Sci U S A* **110**, 1446-1451, doi:10.1073/pnas.1214608110 (2013).

15 Lodato, M. A. *et al.* Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555-559, doi:10.1126/science.aao4426 (2018).

16 Dean, F. B. *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* **99**, 5261-5266, doi:10.1073/pnas.082089499 (2002).

17 Martincorena, I. *et al.* Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348**, 880-886, doi:10.1126/science.aaa6806 (2015).

18 Bae, T. *et al.* Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* **359**, 550-555, doi:10.1126/science.aan8690 (2018).

19 Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* **3**, 246-259, doi:10.1016/j.celrep.2012.12.008 (2013).

20 Alexandrov, L. B. *et al.* Clock-like mutational processes in human somatic cells. *Nat Genet* **47**, 1402-1407, doi:10.1038/ng.3441 (2015).

21 Sniderman, A. D. & Furberg, C. D. Age as a modifiable risk factor for cardiovascular disease. *Lancet* **371**, 1547-1549, doi:10.1016/S0140-6736(08)60313-X (2008).

22 Obas, V. & Vasan, R. S. The aging heart. *Clin Sci (Lond)* **132**, 1367-1382, doi:10.1042/CS20171156 (2018).

23 D'Agostino, R. B., Sr. *et al.* General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* **117**, 743-753, doi:10.1161/CIRCULATIONAHA.107.699579 (2008).

24 Kannel, W. B. & Vasan, R. S. Is age really a non-modifiable cardiovascular risk factor? *Am J Cardiol* **104**, 1307-1310, doi:10.1016/j.amjcard.2009.06.051 (2009).

25 Bui, A. L., Horwich, T. B. & Fonarow, G. C. Epidemiology and risk profile of heart failure. *Nat Rev Cardiol* **8**, 30-41, doi:10.1038/nrcardio.2010.165 (2011).

26 Virani, S. S. *et al.* Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. *Circulation* **141**, e139-e596, doi:10.1161/CIR.0000000000000757 (2020).

27 Zhang, L. *et al.* Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc Natl Acad Sci U S A* **116**, 9014-9019, doi:10.1073/pnas.1902510116 (2019).

28 Weakley, S. M. *et al.* Role of somatic mutations in vascular disease formation. *Expert Rev Mol Diagn* **10**, 173-185, doi:10.1586/erm.10.1 (2010).

29 Reamon-Buettner, S. M. & Borlak, J. Somatic mutations in cardiac malformations. *J Med Genet* **43**, e45, doi:10.1136/jmg.2006.040907 (2006).

30 Patel, A. P. & Natarajan, P. Completing the genetic spectrum influencing coronary artery disease: from germline to somatic variation. *Cardiovasc Res* **115**, 830-843, doi:10.1093/cvr/cvz032 (2019).

31 Strait, J. B. & Lakatta, E. G. Aging-associated cardiovascular changes and their relationship to heart failure. *Heart Fail Clin* **8**, 143-164, doi:10.1016/j.hfc.2011.08.011 (2012).

32 Steinhauser, M. L. & Lee, R. T. Regeneration of the heart. *EMBO Mol Med* **3**, 701-712, doi:10.1002/emmm.201100175 (2011).

33 Sheydina, A., Riordon, D. R. & Boheler, K. R. Molecular mechanisms of cardiomyocyte aging. *Clin Sci (Lond)* **121**, 315-329, doi:10.1042/CS20110115 (2011).

34 Comai, L. The advantages and disadvantages of being polyploid. *Nat Rev Genet* **6**, 836-846, doi:10.1038/nrg1711 (2005).

35 te Beest, M. *et al.* The more the better? The role of polyploidy in facilitating plant invasions. *Ann Bot* **109**, 19-45, doi:10.1093/aob/mcr277 (2012).

36 Gan, P., Patterson, M. & Sucov, H. M. Cardiomyocyte Polyploidy and Implications for Heart Regeneration. *Annu Rev Physiol* **82**, 45-61, doi:10.1146/annurev-physiol-021119-034618 (2020).

37 Fox, D. T. & Duronio, R. J. Endoreplication and polyploidy: insights into development and disease. *Development* **140**, 3-12, doi:10.1242/dev.080531 (2013).

38 Orr-Weaver, T. L. When bigger is better: the role of polyploidy in organogenesis. *Trends Genet* **31**, 307-315, doi:10.1016/j.tig.2015.03.011 (2015).

39 Sawamiphak, S., Kontarakis, Z., Filosa, A., Reischauer, S. & Stainier, D. Y. R. Transient cardiomyocyte fusion regulates cardiac development in zebrafish. *Nat Commun* **8**, 1525, doi:10.1038/s41467-017-01555-8 (2017).

40 Alvarez-Dolado, M. *et al.* Fusion of bone-marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes. *Nature* **425**, 968-973, doi:10.1038/nature02069 (2003).

41 Ali, S. R., Menendez-Montes, I., Warshaw, J., Xiao, F. & Sadek, H. A. Homotypic Fusion Generates Multinucleated Cardiomyocytes in the Murine Heart. *Circulation* **141**, 1940-1942, doi:10.1161/CIRCULATIONAHA.119.043530 (2020).

42 Zeng, B., Tong, S., Ren, X. & Xia, H. Cardiac cell proliferation assessed by EdU, a novel analysis of cardiac regeneration. *Cytotechnology* **68**, 763-770, doi:10.1007/s10616-014-9827-8 (2016).

43 Patterson, M. *et al.* Frequency of mononuclear diploid cardiomyocytes underlies natural variation in heart regeneration. *Nat Genet* **49**, 1346-1353, doi:10.1038/ng.3929 (2017).

44 Bergmann, O. & Jovinge, S. Isolation of cardiomyocyte nuclei from post-mortem tissue. *J Vis Exp*, doi:10.3791/4205 (2012).

45 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 U S A* **117**, 13886-13895, doi:10.1073/pnas.2006163117 (2020).

46 Bohrson, C. L. *et al.* Linked-read analysis identifies mutations in single-cell DNA-sequencing data. *Nat Genet* **51**, 749-754, doi:10.1038/s41588-019-0366-2 (2019).

47 Petljak, M. *et al.* Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell* **176**, 1282-1294 e1220, doi:10.1016/j.cell.2019.02.012 (2019).

48 Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic acids research* **47**, D941-D947, doi:10.1093/nar/gky1015 (2019).

49 Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nature reviews. Genetics* **15**, 585-598, doi:10.1038/nrg3729 (2014).

50 Waters, T. R. & Swann, P. F. Thymine-DNA glycosylase and G to A transition mutations at CpG sites. *Mutat Res* **462**, 137-147, doi:10.1016/s1383-5742(00)00031-4 (2000).

51 De Bont, R. & van Larebeke, N. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**, 169-185, doi:10.1093/mutage/geh025 (2004).

52 Viel, A. *et al.* A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. *EBioMedicine* **20**, 39-49, doi:10.1016/j.ebiom.2017.04.022 (2017).

53 Cadet, J. & Davies, K. J. A. Oxidative DNA damage & repair: An introduction. *Free Radic Biol Med* **107**, 2-12, doi:10.1016/j.freeradbiomed.2017.03.030 (2017).

54 Taverne, Y. J., Bogers, A. J., Duncker, D. J. & Merkus, D. Reactive oxygen species and the cardiovascular system. *Oxid Med Cell Longev* **2013**, 862423, doi:10.1155/2013/862423 (2013).

55 Aguiar, P. H. *et al.* Oxidative stress and DNA lesions: the role of 8-oxoguanine lesions in Trypanosoma cruzi cell viability. *PLoS Negl Trop Dis* **7**, e2279, doi:10.1371/journal.pntd.0002279 (2013).

56 Alexandrov, L. B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101, doi:10.1038/s41586-020-1943-3 (2020).

57 Drost, J. *et al.* Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* **358**, 234-238, doi:10.1126/science.aao3130 (2017).

58 Liu, D., Keijzers, G. & Rasmussen, L. J. DNA mismatch repair and its many roles in eukaryotic cells. *Mutat Res* **773**, 174-187, doi:10.1016/j.mrrev.2017.07.001 (2017).

59 Kunkel, T. A. & Erie, D. A. Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu Rev Genet* **49**, 291-313, doi:10.1146/annurev-genet-112414-054722 (2015).

60 Nemeth, E. *et al.* Two main mutational processes operate in the absence of DNA mismatch repair. *DNA Repair (Amst)* **89**, 102827, doi:10.1016/j.dnarep.2020.102827 (2020).

61 Kozmin, S. G., Sedletska, Y., Reynaud-Angelin, A., Gasparutto, D. & Sage, E. The formation of double-strand breaks at multiply damaged sites is driven by the kinetics of excision/incision at base damage in eukaryotic cells. *Nucleic Acids Res* **37**, 1767-1777, doi:10.1093/nar/gkp010 (2009).

62 Iyama, T. & Wilson, D. M., 3rd. DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst)* **12**, 620-636, doi:10.1016/j.dnarep.2013.04.015 (2013).

63 Consortium, G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660, doi:10.1126/science.1262110 (2015).

64 Campbell, B. B. *et al.* Comprehensive Analysis of Hypermutation in Human Cancer. *Cell* **171**, 1042-1056 e1010, doi:10.1016/j.cell.2017.09.048 (2017).

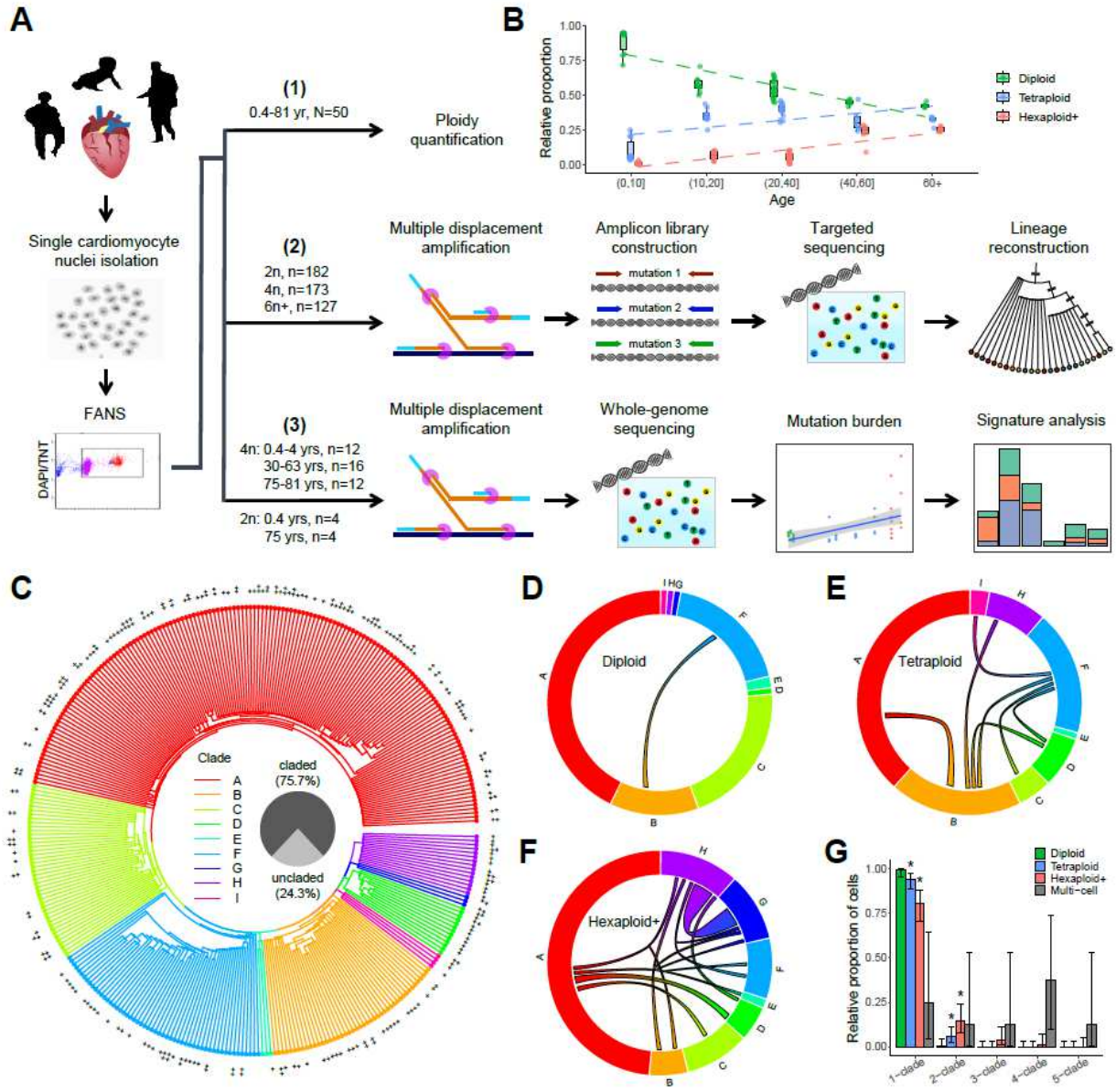


Figure 1. Lineage reconstruction of human cardiomyocytes via single-cell sequencing based on ploidy. (A) Schematic of approach. Nuclei are isolated from frozen postmortem heart tissues with varied ages, and sorted based on ploidy content by FANS. Sorted nuclei are subjected to ploidy quantification (upper), or amplified by Φ29 polymerase-mediated MDA for targeted sequencing for cell lineage analysis (middle) or whole-genome sequencing for sSNV burden and signature analysis (lower) after. The numbers of examined individuals (N) and nuclei (n) are marked. (B) Increased ploidy in cardiomyocyte with age. Quantification of cardiomyocyte ploidy, from normal postmortem heart of 50 individuals ranging from 0.4 - 82.7 years age,

indicating increased ploidy in cardiomyocyte with age. At birth most cardiomyocytes are diploid with only ~5% polyploid cardiomyocytes. The polyploidy proportion increases to ~50% by age 60. Boxplot denotes median with 25% and 75% quartile. (C) Lineage map of 340 single-clade human cardiomyocytes based on panel sequencing of 253 clonal sSNVs. Cardiomyocytes are placed into nine distinct clades defined by one or more clade-specific sSNVs. Tetraploid nuclei and hexaploid or higher (hexaploid+) nuclei are labeled by + and ++, respectively. (D-F) Cardiomyocytes generated by fusion between cells from multiple clades. The connecting arch between clades indicates nuclei containing sSNVs from more than one clade, and the thickness of arch indicates the number of nuclei. All but one diploid nuclei (D) belong to single clade, suggesting the low double-sorting rate, whereas tetraploid (E) and hexaploid+ nuclei (F) have much larger fractions of multi-clade cells generated by fusion. (G) Proportion of cardiomyocytes with different clade numbers. Tetraploid and hexaploid+ cardiomyocytes show higher proportion of multi-clade cells than diploid cells (two-tailed proportion test, asterisk, $p < 0.05$,), suggesting fusion as a mechanism for polyploidization of human cardiomyocytes. Error bar: 95% confidence interval.

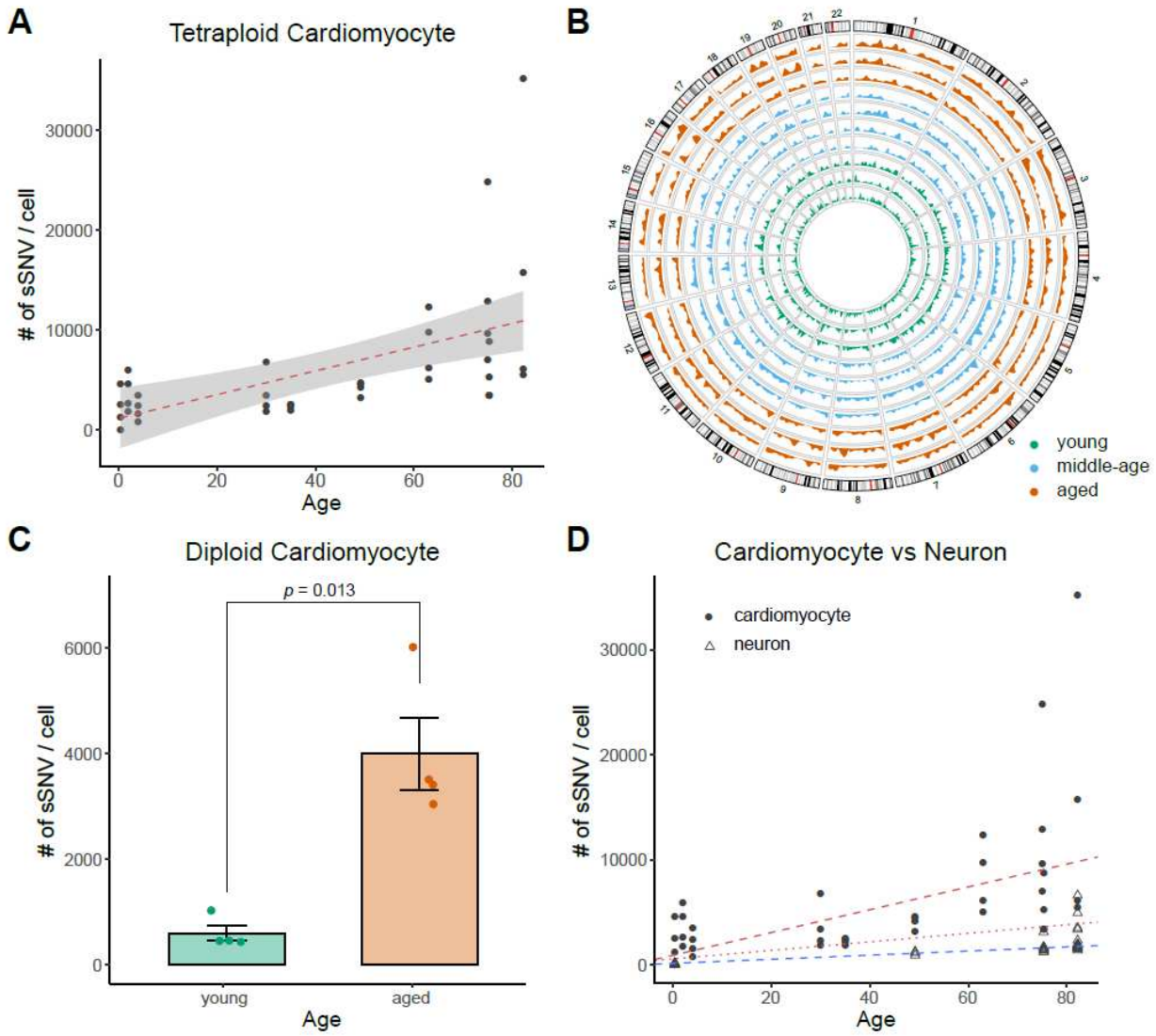


Figure 2. sSNV burden in single human cardiomyocytes detected by whole-genome sequencing. (A) Estimated genome-wide sSNV burden of tetraploid cardiomyocytes. Tetraploid cardiomyocytes showed an increased number of sSNVs with increased age (~2,600 in young vs ~11,500 in aged). Single tetraploid cardiac nuclei were obtained from normal postmortem heart of individuals ranging from 0.4 to 82.7 years of age. (B) Genome-wide distribution of SNVs. Autosomal sSNV density in 10Mb genomic window was plotted for each individual. Green, blue, and red represents young, middle-age, and aged hearts, respectively. (C) Estimated genome-wide sSNV burden of diploid cardiomyocytes. Aged diploid cardiomyocytes showed ~7 fold increase in sSNV burden compared to young cardiomyocytes (~600 in young vs ~4,000 in aged) Error bar: standard error of the mean. Single diploid cardiac nuclei were obtained from normal postmortem heart of a young (0.4

yrs.) and an aged (75 yrs.) individual. (D) Age-dependent increases of sSNV burden in cardiomyocytes *vs* neurons. Human cardiomyocytes demonstrate a significantly faster sSNVs accumulation than neurons (~2,600 vs ~100 in young and ~11,500 vs ~2,000 in aged). Red and blue dotted lines indicate the regression lines of cardiomyocytes (thick for tetraploid and thin for diploid) and neurons, respectively.

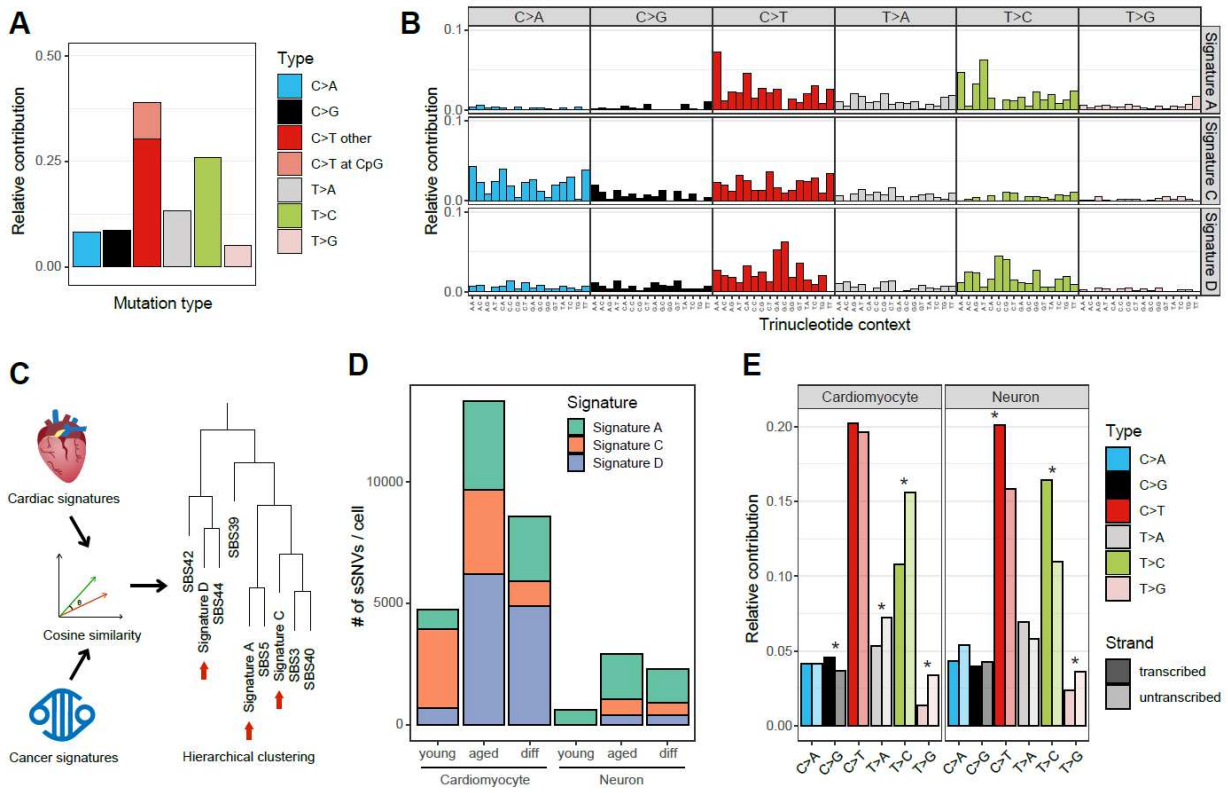


Figure 3. Signature analysis of sSNV reveals mutational process in cardiomyocytes during aging. (A) Substitution type for the age-accumulated sSNVs (the "net increase" of sSNVs between young and aged cardiomyocytes). C>T and T>C substitutions are predominant. (B) Cardiac mutational signatures identified by non-negative matrix factorization based on the trinucleotide context of sSNVs. Each signature is displayed according to the 96 trinucleotide contexts, defined by the six substitution types and sequence context immediately 5' and 3' to the mutated base. Although both Signatures A and D predominate with C>T and T>C substitutions, they differ in trinucleotide contexts at C>T and T>C substitutions. (C) Hierarchical clustering between cardiac and COSMIC cancer signatures. Cosine similarity of 96 trinucleotide contexts was used to measure the pairwise distance between signatures. Signature A resembles SBS5 and Signature D resembles SBS44. (D) Signature-specific sSNV burden in human cardiomyocytes and neurons. The "diff" group represents the age-accumulated sSNVs in cardiomyocytes or neurons. Signature D specifically accumulated in aged cardiomyocytes, whereas Signature A accumulated in both cardiomyocytes and neurons. (E) Transcription bias of age-accumulated sSNVs. Asterisk mark significant difference between transcribed and untranscribed strands ($p < 0.05$, Poisson test). Notably, T>C substitution is enriched in the untranscribed strand

in cardiomyocytes whereas it is enriched in the transcribed strand in neurons, suggesting different mutation mechanisms in these two nondividing cells.

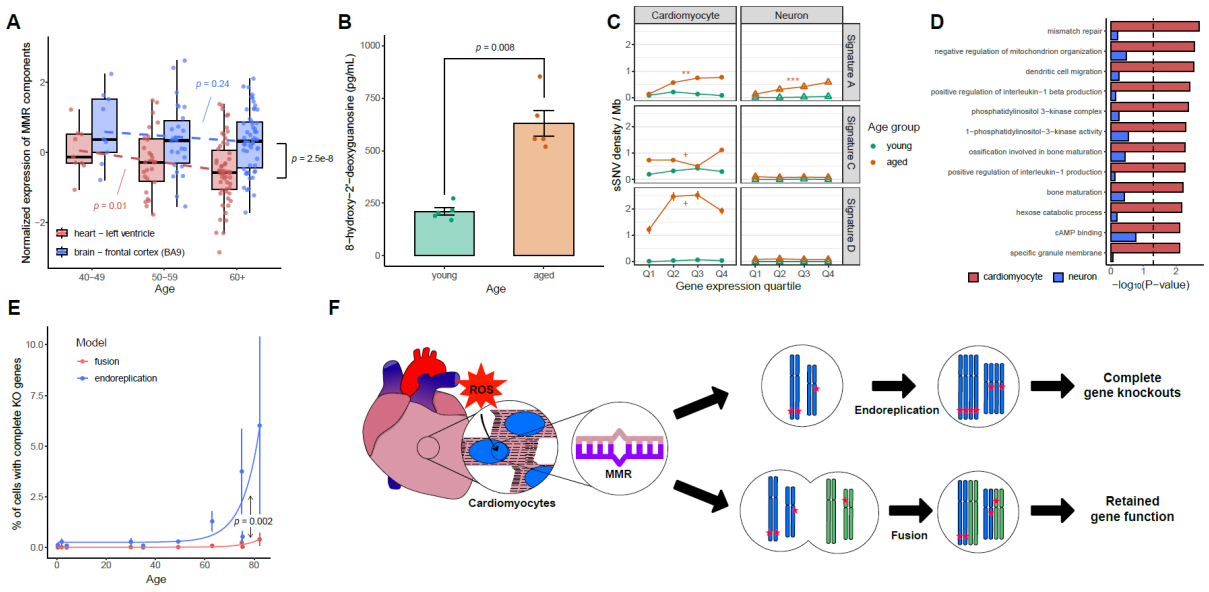


Figure 4. Potential mechanism of sSNV formation and their impact on cardiomyocytes genome during aging. (A) mRNA expression level of MMR complex components in human heart and brain samples of GTEx. Key MMR genes are significantly down-regulated with increased age in heart but not in brain. Boxplot denotes median with 25% and 75% quartiles. (B) Quantification assay of 8-hydroxy-2-deoxyguanosine from human cardiomyocytes. Aged heart shows significantly higher oxidative stress than young heart. Error bar: standard error of the mean. (C) Association of sSNVs with gene expression level. Signature A accumulated more sSNVs in highly expressed genes in both aged cardiomyocytes and neurons. Signatures C and D are dominant in cardiomyocytes without strong correlation with gene expression level. Error bar: 95% confidence interval. +: $p < 0.1$; **: $p < 0.01$; ***: $p < 0.001$. (D) Cardiomyocyte-specifically enriched gene ontology (GO) categories. GO terms with enrichment in cardiomyocytes but not in neurons are listed (FDR-adjusted $p < 0.05$, permutation test). Cardiomyocyte sSNVs are enriched in MMR pathways and pathways involved in metabolism and kinase signaling. (E) Prediction model for effect of sSNVs on abundance of "knockout" tetraploid cardiomyocytes. The percentage of tetraploid cardiomyocytes containing at least one tetra-allelic knock-out gene due to deleterious exonic sSNVs was estimated using fusion and endo-replication models. The fusion model demonstrates a significantly lower "knockout" rate than the endo-replication model, especially in aged heart, suggesting the protective effect of fusion against deleterious sSNV accumulation. Error bar: 95% confidence interval. (F) Mechanism of sSNV occurrence in heart and adaptation of cardiomyocyte to a fusion

mediated polyploid genome. Cardiomyocytes with increased age show increased level oxidative stress (ROS), which results in a mismatched pairing in the genome. Defective MMR in aged cardiomyocytes might not function effectively to repair this increased load of mismatches and lead to fixed sSNVs. Cardiomyocytes with higher ploidy, generated by fusion, can better tolerate the deleterious effect of these mutational burdens.

Materials and Methods:

Human tissues DNA sample preparation and isolation of cardiomyocyte nuclei:

All human tissues have been obtained from the NIH NeuroBioBank at the University of Maryland. Once we received the tissue from the BioBank, DNA degradation evaluation was performed by isolating DNA from the tissue and performing gel electrophoresis as well as a Genomic DNA Screen Tape Station. Tissues with fragmented DNA were not selected for further studies. We have sequenced 40 tetraploid and 8 diploid single cardiac nuclei from 9 individuals with age from 0.4 to 82 years, including 3 young (<4 yrs.), 3 middle-age (30-63 yrs.) and 3 aged (>75 yrs.) individuals (Supplementary Table 3). Samples were processed according to a standardized protocol under the supervision of the NIH NeuroBioBank ethics guidelines. This study was approved by the Boston Children's Hospital IRB. Bulk DNA was extracted using the QIAamp DNA Mini kit with RNase A treatment. Cardiac nuclei were isolated utilizing a density sedimentation protocol[1]. Subsequently, single cardiac nuclei were isolated using FANS-based cardiac troponin T or PCM1 nuclear staining and DAPI intensity. Cardiomyocytes are the only tetraploid cell in cardiac tissue, so ploidy is a convenient way to purify them. However, to rule out any possibility that tetraploid cells accumulate mutations in different ways from diploid cells, or that the amplification, sequencing and calling process performs differently in tetraploid cells, we carried out replicate analysis of diploid cells by isolating diploid cardiomyocytes using cardiomyocyte-specific markers (cardiac troponin T and PCM1) and performing MDA and single-cell WGS. The genomes of each single nucleus were amplified using multiple displacement amplification (MDA)[2].

Library preparation and whole-genome sequencing:

Low coverage library preparations were carried out according to manufacturer's instructions for QIAseq FX single-cell DNA Kit. Paired-end barcoded WGS libraries were prepared with the Illumina TruSeq Nano LT sample preparation kit. Paired-end sequencing (150bp×2) was performed on a HiSeq X10 instrument. Single-cell WGS library preparations and sequencing were done at Macrogen Genomics. MDA-amplified DNA libraries were generated using Illumina TruSeq Kits and sequenced by Macrogen Genomics Illumina HiSeq X10 instruments.

Targeted sequencing of somatic SNVs:

Amplicon primers for 253 candidates clonal sSNVs were designed for the amplicon size of 275 bp with Ion AmpliSeq Designer from Thermo Fisher Scientific. MDA-amplified cardiac nuclei were diluted 10 folds and amplicon-based enrichment was performed with pooled amplicon primers. Illumina sequencing library preparation for highly multiplexed target capture was done according to Meyer R Kircher's protocol as described elsewhere[3] Briefly Amplicons were generated using a high-fidelity polymerase and then were purified using a magnetic bead capture kit (Ampure; Agencourt) and quantified using a fluorometric kit (QuantIT PicoGreen; Invitrogen). The purified amplicons were then pooled in equimolar concentrations and were sequenced on Illumina HiSeq X10 with an average depth of >7000X (Extended Data Fig. 1).

Read alignment and post-processing:

Reads generated from single-cell WGS and targeted sequencing were aligned against the GRCh37 human reference genome by BWA (ver 0.7.15)[4] with default parameters. For single-cell WGS data, duplicate reads were masked by Mark Duplicate of Picard (ver 2.8) and then post-processed with local realignment around indels and base quality score recalibration using Genome Analysis Toolkit (GATK) (ver 3.5)[5]. For single-cell targeted sequencing data: 1) read pairs with overlapping tails were merged into single consensus reads using USEARCH (ver 11.0.667)[6], 2) error-prone reads with gap alignment or >4 mismatches were removed, and 3) the 5bp regions at both ends of the consensus reads were masked due to its vulnerability to systematic amplification and sequencing errors.

Somatic SNV genotyping from single-cell targeted sequencing:

The single-cell genotype likelihood was calculated for each of the 253 sSNV sites using the Bayesian model of MosaicHunter[7,8] with three possible genotypes: homozygous for reference allele (ref-hom), heterozygous (het), or homozygous for the alternative allele (alt-hom). A minimal depth of 30X was required for confident single-cell genotyping. In each single-cell an sSNV was called if 1) the likelihood of ref-hom was <0.5; 2) ≥5 reads supported the mutant allele; and 3) the mutant allele fraction was ≥0.05. The sSNV sites that were absent in all sequenced single-cells were excluded from subsequent analyses.

Lineage reconstruction using clonal somatic SNVs:

To reconstruct the lineage tree of 482 cardiomyocytes subjected to single-cell targeted sequencing, we calculated the pairwise genetic similarity (*S*) as the cosine similarity of binarized sSNV profiles between two cardiomyocytes, where called sSNVs were

considered as “1” and absent sSNVs were considered as “0”. Unsupervised clustering was then performed using the UPGMA (unweighted pair group method with arithmetic mean) method with the distance defined as 1-*S*. The clustered cardiomyocytes were grouped into 9 distinct nested clades (clades A-I in Figure 1C), where each clade was defined by one or more clade-specific sSNVs (*i.e.*, sSNVs present in only one clade but absent in other clades). For 93.2% (340/365) of the claded cardiomyocytes, they only harbored sSNVs from a single clade (Supplementary Table 1). In comparison, 25 cardiomyocytes (1 diploid, 8 tetraploid, and 16 hexaploid+) contained sSNVs belonging to more than one clades, suggesting their multi-clade origins that could be explained by fusion events between cells from different clades.

Sensitivity correction for ploidy proportion:

In our sSNV-based lineage reconstruction, the detection sensitivity of cardiomyocytes derived from potential fusion events was affected by 1) whether the clade-informative sSNVs were successfully MDA-amplified and sequenced in targeted sequencing; 2) whether the origin cells were from two or more genetically distinct clades or the same clade because the latter case was not distinguishable from cardiomyocytes without fusion. To address this, we developed a mathematical model to estimate the detection sensitivity in profiled diploid, tetraploid, and hexaploid+ cardiomyocyte populations, separately. For simplicity, we only considered the fusion events involving two cells because they dominated 84% (21/25) of the detected multi-clade cells. First, the probability of capturing any clade-*i* sSNVs (r_i) was modeled below,

$$r_i = \sum_{n=1}^{N} S_{in}(1-(AD+LD)^n)$$

where N denotes the number of sSNV in clade-i, and S_{in} denotes the proportion of cells in clade-i carrying n clade-informative sSNV ($1 \leq n \leq N$). *AD* and *LD* denote the overall allelic and locus dropout rates estimated from the genotyping profile of targeted sSNVs, respectively. *LD* was calculated as the proportion of sites with <30X depth, and *AD* was calculated as below with the assumption that reference and alternative alleles had an equal dropout chance and all alt-hom genotypes of sSNVs in single-cells were resulted from allelic dropout of the reference allele,

$$AD = \frac{2P_{alt\text{-}hom}}{(P_{het} + 2P_{alt\text{-}hom})}$$

where P_{het} and $P_{alt\text{-}hom}$ are the proportion of clade-informative sSNVs with heterozygous and alt-hom genotypes, respectively.

Next, we modeled the overall probability of detecting a fusion event by summing up all possible combinations of clades i and j, after correcting for the clade-specific sensitivity due to allelic and locus dropout described previously,

$$sensitivity = \frac{\sum_i \sum_j p_i r_i p_j r_j c_{ij}}{\sum_i \sum_j p_i p_j}$$

$$c_{ij} = \begin{cases} 0 & i = j \\ 1 & i \neq j \end{cases}$$

where p_i and p_j denotes the proportion of cells belonging to clade-i and j, and c_{ij} is a binary value to model that a fusion event can only be detected when two origin cells are from different clades. As a result, the detection sensitivity for fusion cells was estimated

as 0.21, 0.46, and 0.3 in diploid, tetraploid, and hexaploid+ cardiomyocytes, respectively (Supplementary Table 2).

Measuring the evenness of genome amplification in single cells:

We measured the evenness of genome amplification in single-cells using two metrics: median absolute pairwise difference (MAPD) and coefficient of variation (CoV). MAPD is the median value of absolute differences between the copy number ratios of neighboring bins with variable length where bins were divided to have the same number of uniquely mapping reads. CoV is the measure of variance of bin-wise copy number ratios, calculated by taking the ratio of their standard deviation to the mean. Both higher MAPD and CoV scores represent greater unevenness of single cell genome amplification. Binning, GC normalization, segmentation, and copy estimation were performed following the previous single-cell copy number analysis protocol[9], to obtain the copy number ratio per bin and calculate MAPD and CoV scores.

Estimation of amplification dropout rates:

Germline heterozygous SNVs were identified from bulk WGS data using GATK[5], and only those reported by the 1000 Genomes Project[10] were considered subsequently as high-confidence variants. For each single-cell, a germline heterozygous SNV was considered as "locus-dropout" if the total coverage in single-cell WGS is less than 5X and considered as "allele-dropout" if the number of reads supporting either a reference or a mutant allele is less than two, and the genome-wide locus- and allele-dropout rates were

then calculated as the proportions of dropout sites among all germline heterozygous SNVs.

Somatic SNV calling from single-cell WGS data:

We performed phasing-based linked read analysis using the LiRA method (ver 2018Feb)[11] to identify sSNVs in single-cells using ~30X WGS data of non-heart tissue from the same individual as bulk germline controls. Initial somatic and germline SNVs were called using GATK's HaplotypeCaller and germline SNVs were further phased by Shapeit 2 (ver 904)[12]. LiRA distinguishes true somatic mutations from base-calling or amplification errors by leveraging the linkage information between the somatic candidate and adjacent phased germline mutations. Genome-wide sSNV burden was further calculated by correcting for the detection power of phasable regions estimated from germline SNVs. We considered sSNVs in autosomes only to avoid potential detection bias in sex chromosomes between different genders.

Genome-wide sSNV burden correction for tetraploid cells:

Each tetraploid cardiomyocyte contains two sets of diploid genomes (*i.e.*, four haplotypes). Theoretically, somatic mutations present in one out of the four haplotypes can be called by LiRA only when the reads violating complete linkage were lost due to allelic or locus dropout (Extended Data Fig. 8). In comparison, germline mutations can be called in the same way as in the diploid cells. Therefore, the LiRA-estimated power from germline SNVs should be corrected by cell-specific allelic or locus dropout rates before applying to the count of identified sSNVs in tetraploid cells.

By definition, the dropout status of a site in a tetraploid cell (S_{4n}) was determined by the status of its two diploid origins as below,

$$S_{4n} = \begin{pmatrix} LD & mut\text{-}AD & ref\text{-}AD & normal \\ mut\text{-}AD & mut\text{-}AD & normal & normal \\ ref\text{-}AD & normal & ref\text{-}AD & normal \\ normal & normal & normal & normal \end{pmatrix}$$

where each row and column denote different dropout status of two original diploid cells: locus dropout (LD), allelic dropout for the mutant allele (mut-AD), allelic dropout for the reference allele (ref-AD), and normal. Using simulated tetraploid cells by *in silico* mixture of single-cell WGS data from two diploid cells from the same individual, we observed that the locus dropout rate in simulated tetraploid cells generally equals to the product of the dropout rates in the two original diploid cells, suggesting independence of dropout occurrences across the genome between MDA-amplified cells.

Cell-specific allelic and locus dropout rates were estimated for each tetraploid cell, and then the dropout rate for its two diploid origins was calculated from the above matrix under the assumption of equal rates between the two diploid origins. The count and burden of LiRA-called sSNVs were further adjusted by the locus and allelic dropout rate of the corresponding diploid origins and the sensitivity loss due to the decrease of per-haplotype sequencing depths from diploid to tetraploid cells.

Mutational signature analysis:

Mutational signatures were *de novo* decomposed by the non-negative matrix factorization (NMF)-based mutational signature framework[13] using Mutational Patterns (ver 1.8.0)[14] , using the 96 trinucleotide contexts of sSNVs detected from tetraploid cardiomyocytes in this study as well as non-disease neurons that were previously studied[15]. We estimated signature stability and reconstruction error and found that four signatures best fit the observed sSNV profiles (Extended Data Fig. 3). We then compared our *de novo* signatures (Signatures N1, N2, N3, and N4) with previously reported signatures in neurons (Signatures A, B, and C)[15] and signatures potentially resulted from single-cell artifacts (SBS scE and scF)[16]. As shown in Extended Data Fig. 4, Signature N1, N2, and N3 resemble Signature B (as well as SBS scF), A, and C, respectively, whereas Signature N4 did not show high similarity to any of these signatures, suggesting a potential cardiac-specific signature (renamed as Signature D).

Considering the evidence that SBS scF (highly similar to Signature N1/B) represents potential single-cell artifacts[16], we removed the contributions of Signature N1/B in the estimation of genome-wide sSNV burden per cell. Specifically, we decomposed the sSNV profile of each single cell using Signatures N2/A, N3/C, N4/D as well as SBS scE and scF (Signature N1/B was not included because it was nearly identical to SBS scF) using MutationalPatterns[14]. Then we calculated the number of sSNVs derived from SBS scF by multiplying the LiRA-estimated sSNV burden and the cell-specific weight of SBS scF and then subtracted the sSNV count from SBS scF in subsequent burden analyses.

Mutation spectrum and strand bias analysis:

The LiRA-identified sSNVs were grouped into three categories according to the age of their cell donor: young (<4 yrs.), middle-age (30-63 yrs.), and aged (>75 yrs.), and then the mutation spectrum and strand bias were calculated for each age category. The transcriptional strands of genic sSNVs were assigned based on the UCSC TxDb annotations by MutationalPatterns[14], where mutated bases ("C" or "T") on the same strand as the gene direction were categorized as "untranscribed" and on the opposite strand as "transcribed". To characterize sSNV accumulation during aging, we further estimated the mutation spectrum and strand bias for the net increase of sSNVs between young and aged categories. Specifically, we first measured the absolute sSNV count for each mutation type by multiplying its proportion and the average genome-wide sSNV burden for each age category, and then subtracted the sSNV count for each mutation type between young and aged categories. Statistical significance of strand bias was determined by the Poisson test.

Gene expression analysis:

The expression matrix for left ventricle of heart and frontal cortex (BA9) of brain was downloaded from GTEx[17]. The per-gene expression value was normalized for each individual after controlling for age and gender using DESeq2 (ver 1.24.0)[18]. To study the age-dependent changes in MMR activity in the heart and brain, we extracted the expression levels of *MLH1*, *MLH3*, *MSH2*, *MSH3*, *MSH6*, *PMS1* and *PMS2*, which encode the core components of the MMR complex. Individuals with both heart and brain expression profiles were binned according to their ages. Individuals with age <40 years (n=6) were excluded due to the small sample size. For the remaining 178 individuals, we

calculated the average expression levels of the seven MMR genes in heart and brain samples, separately, and tested their association with age using the linear regression model.

To investigate the relationship between somatic mutation and gene expression, we assigned genes into four quartiles based on their average expression values in heart or brain samples across all GTEx individuals. Cardiac and neuronal sSNV densities were calculated for each quartile of genes, after normalizing by gene length and per-cell sSNV detection power. The standard deviation of sSNV density was estimated using a permutation test, in which the quartile classification of genes was randomly shuffled and the permuted sSNV densities were calculated for 1000 rounds. We further performed a NMF-based mutational signature decomposition for sSNVs located in each quartile of genes, to estimate the relative contributions of Signature A, Signature C, Signature D, SBS scE, and SBS scF for each quartile. The sSNV density for each signature was calculated by multiplying the overall sSNV density by the signature contribution. We also performed the above analysis by using the expression profiles from aged individuals (>75 yrs.) only and observed robust results.

Functional enrichment analysis:

Gene Ontology (GO) functional enrichment analysis was performed using GOseq (ver 1.34.1)[19]. We assigned a binary value "0" or "1" to each RefSeq gene according to whether any sSNV was present in the gene in any single cell and built the sSNV-gene table for cardiac and neuronal sSNVs separately. GOseq uses the Wallenius

approximation method to test the enrichment of sSNVs for each GO term, after applying a probability weighting function to control for potential bias from gene length. Genes without any GO annotation were ignored when calculating the total gene count. GO terms with less than 5 member genes with sSNVs were excluded to avoid ascertainment bias. GO terms with more than 1000 member genes were also exclueded.

To identify GO terms that were specifically enriched in cardiomyocytes but not in neurons, we performed a permutation test among all GO terms with $p < 0.01$ for either cardiac or neuronal sSNVs. For each permutated GO term, we compared the observed rank difference in GOseq's *p* between cardiac and neuronal sSNVs against the expected null distribution, which was estimated by 1000 rounds of random shuffling of the sSNV-gene tables. The FDR method was applied for correcting multiple hypothesis testing.

Measurement of oxidative stress:

The level of 8-hydroxy-2′deoxyguanosine and 8-hydroxyguanosine was measured in 250 ng total nucleic acids extracted from left ventricular cardiomyocytes using a competitive enzyme-linked immunosorbent assay kit (#589320, Cayman Chemical) according to the manufacturer's instructions. The samples were incubated for 1 hour with a monoclonal antibody against 8-hydroxy-2′deoxyguanosine in a micro titer plate pre-coated with 8-hydroxy-2′deoxyguanosine and 8-hydroxyguanosine. The final color was developed by the addition of 3,3,5,5-tetramethylbenzidine, and absorbance was measured at 450 nm. The samples were diluted 1:50 with EIA buffer before assaying.

Modeling accumulation of gene knock-outs in cardiomyocytes:

Accumulation of exonic, deleterious "gene knockout" (KO) mutations might be detrimental for proper cell function. These mutations can be "biallelic" in the case of diploid cells, or "quadallelic" in the case of tetraploid cells. The number of sSNVs identified in this study were used to estimate the accumulation of gene KOs in single cells, using an extension of the method described in Lodato et al[20]. To account for genes that are highly dosage sensitive and thus can be haploinsufficient, we included a factor to capture the probability of a mutation landing on an allele of a gene with a high pLI score. The pLI metric measures the probability of loss-of-function intolerance[21], and genes with pLI > 0.90 are considered highly dosage sensitive. These high pLI score genes comprise 17% of all genes. Consequently, the calculation was computed as follows:

$$n = \#\ estimated\ sSNVs\ \times \frac{\text{total \# deleterious variants}}{\text{total \# variants}} \times\ p$$

d_i = {event that gene *i* has at least one mutation}

π_i = {event that gene *i* has a high pLI score}

D = {probability of a gene having a deleterious mutation}

$$\Pr(KO|\pi, D, n) = \pi \times (1 - (1 - D)^n) + (1 - \pi)(1 - e^{-nD})$$

where p is the ploidy factor which captures the probability of obtaining a mutation on the remaining alleles of the gene (e.g., p=0.5 for diploid genomes, and p=0.125 for tetraploid genomes), and n is the expected number of deleterious mutations (i.e. frameshift, splice altering, stop gain). The average was taken across all cells per individual and 95% confidence interval on those point estimates were calculated. Regressions were performed using a log-normal model to capture the non-linear trend of the probability of

obtaining cells with KO genes with age. All calculations were performed using custom R scripts.

Data availability:

All the single-cell WGS data for cardiomyocytes will be deposited in the NCBI SRA.

Code availability:

Code used in this study is available upon reasonable request to the corresponding authors.

Reference:

1 Bergmann, O. & Jovinge, S. Isolation of cardiomyocyte nuclei from post-mortem tissue. *Journal of visualized experiments : JoVE*, doi:10.3791/4205 (2012).

2 Dean, F. B. *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* **99**, 5261-5266, doi:10.1073/pnas.082089499 (2002).

3 Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols* **2010**, pdb prot5448, doi:10.1101/pdb.prot5448 (2010).

4 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

5 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics* **43**, 11 10 11-33, doi:10.1002/0471250953.bi1110s43 (2013).

6 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461, doi:10.1093/bioinformatics/btq461 (2010).

7 Xu, X. *et al.* Amplicon Resequencing Identified Parental Mosaicism for Approximately 10% of "de novo" SCN1A Mutations in Children with Dravet Syndrome. *Hum Mutat* **36**, 861-872, doi:10.1002/humu.22819 (2015).

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, doi:10.1093/nar/gkx024 (2017).

9 Baslan, T. *et al.* Genome-wide copy number analysis of single cells. *Nat Protoc* **7**, 1024-1041, doi:10.1038/nprot.2012.039 (2012).

10 Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073, doi:10.1038/nature09534 (2010).

11 Bohrson, C. L. *et al.* Linked-read analysis identifies mutations in single-cell DNA-sequencing data. *Nat Genet* **51**, 749-754, doi:10.1038/s41588-019-0366-2 (2019).

12 Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nat Methods* **9**, 179-181, doi:10.1038/nmeth.1785 (2011).

13 Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell reports* **3**, 246-259, doi:10.1016/j.celrep.2012.12.008 (2013).

14 Blokzijl, F., Janssen, R., van Boxtel, R. & Cuppen, E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* **10**, 33, doi:10.1186/s13073-018-0539-0 (2018).

15 Lodato, M. A. *et al.* Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555-559, doi:10.1126/science.aao4426 (2018).

16 Petljak, M. *et al.* Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell* **176**, 1282-1294 e1220, doi:10.1016/j.cell.2019.02.012 (2019).

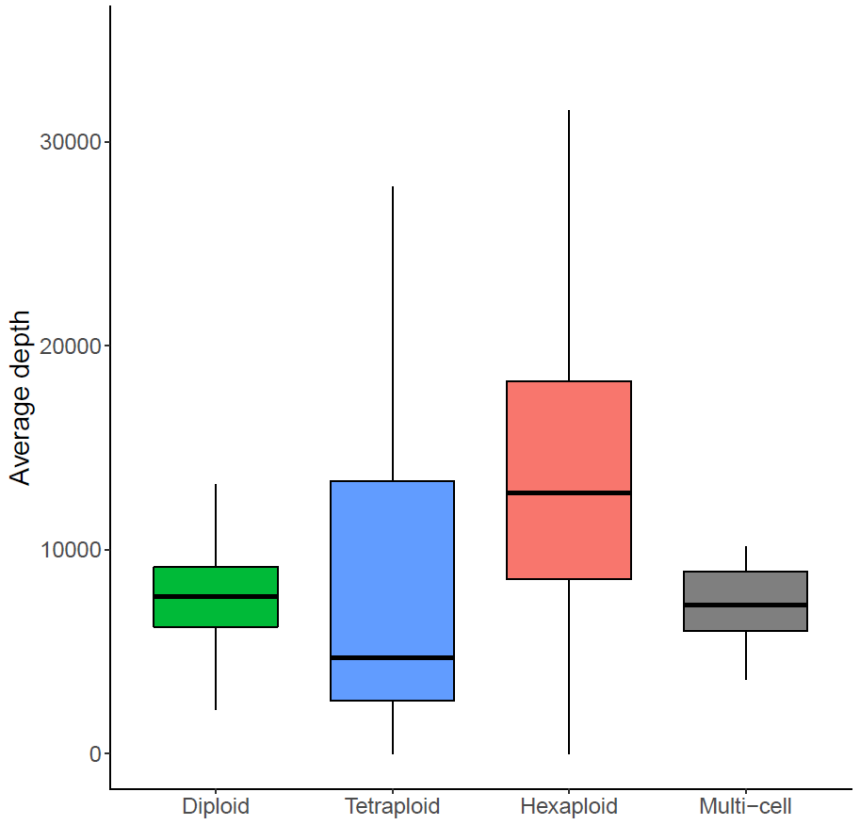
17 Consortium, G. T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213, doi:10.1038/nature24277 (2017).

18 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

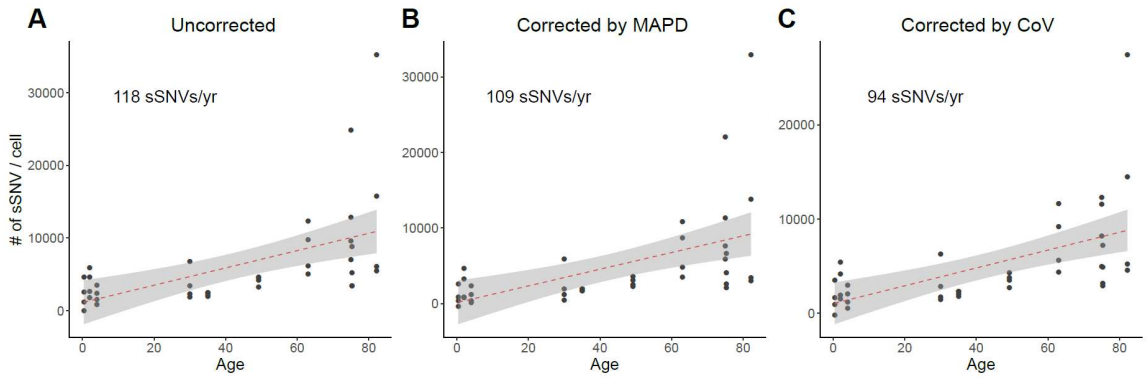
19 Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* **11**, R14, doi:10.1186/gb-2010-11-2-r14 (2010).

20 Lodato, M. A. *et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94-98, doi:10.1126/science.aab1785 (2015).

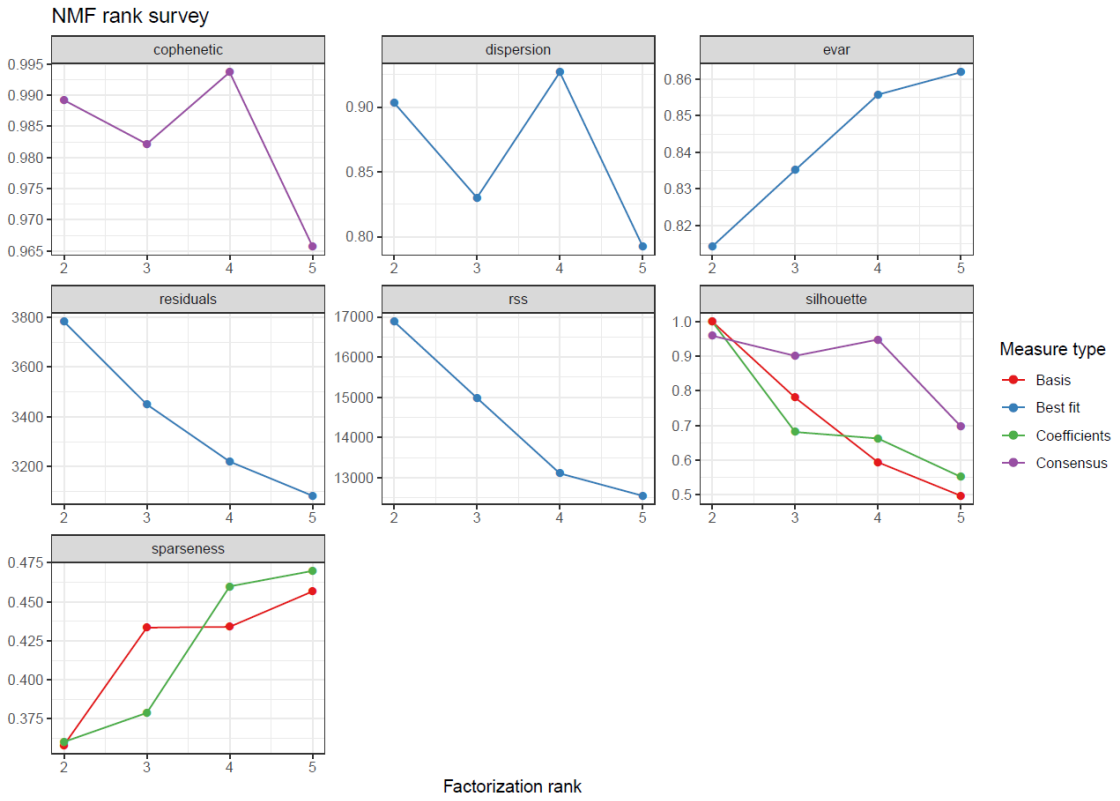
21 Karczewski, K. J. *et al.* The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic acids research* **45**, D840-D845, doi:10.1093/nar/gkw971 (2017).



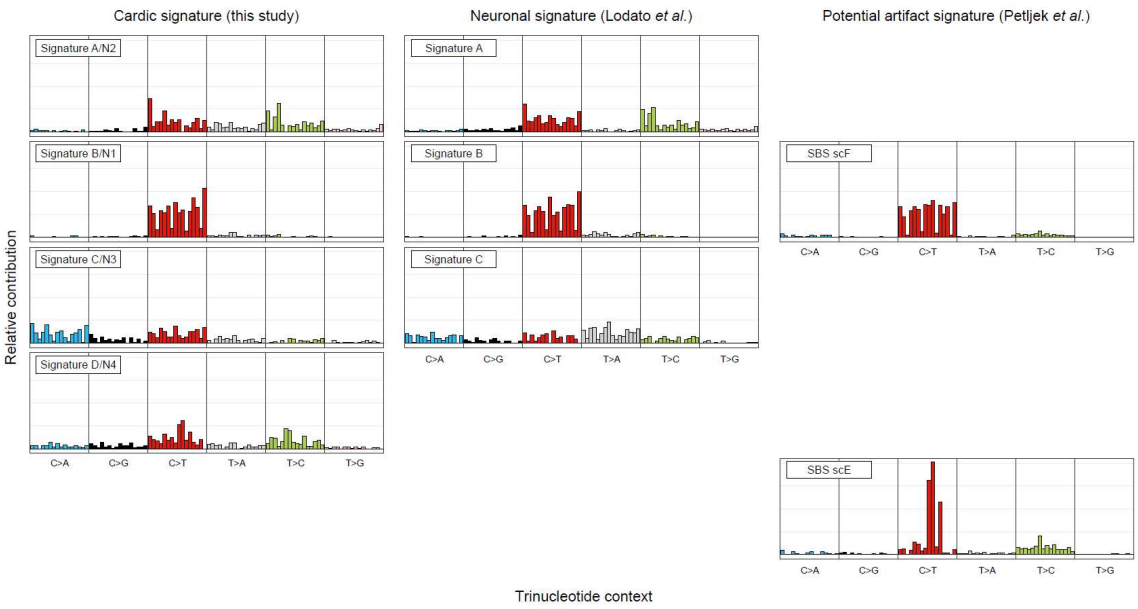
Extended Data Fig. 1. Average depth for cardiac nuclei amplicon panel sequencing. For targeted panel sequencing of 253 clonal sSNVs, 482 MDA amplified cardiac nuclei were sequenced with an average depth of >7000X. Boxplot denotes median with 25% and 75% quantile.



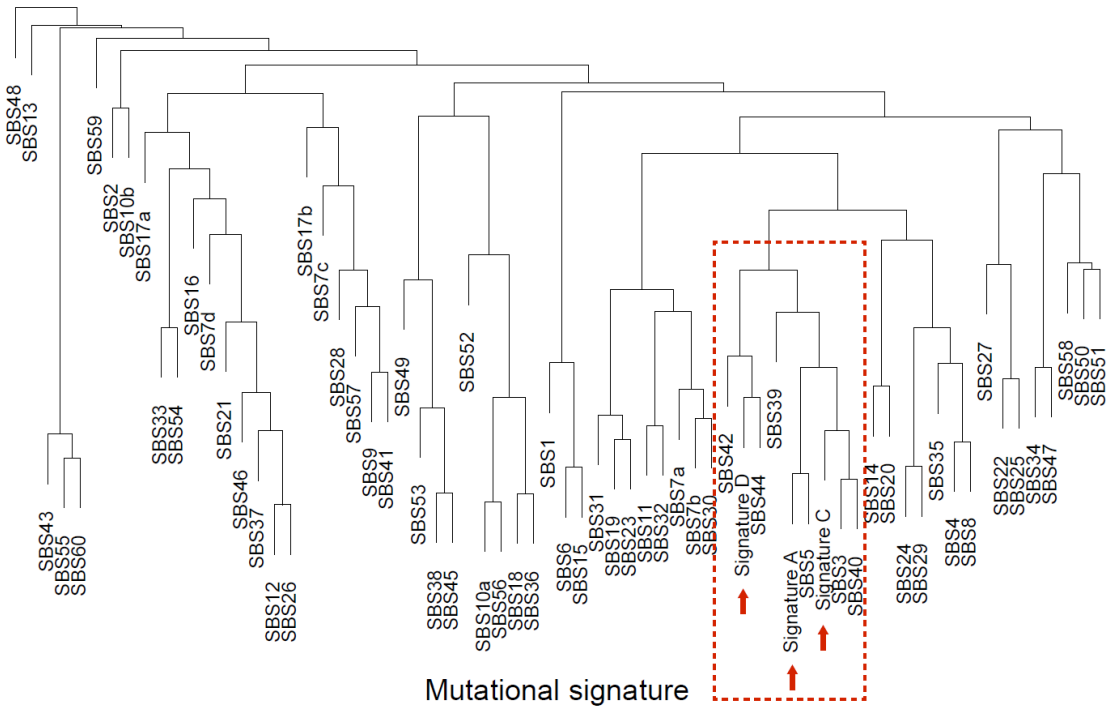
Extended Data Fig. 2. Evaluation of sSNVs burden in cardiomyocytes with the consideration of amplification evenness. (A) Uncorrected genome-wide sSNV burden of tetraploid cardiomyocytes. (B-C) Genome-wide sSNV burden of tetraploid cardiomyocytes after correcting for two metrics about amplification evenness, MAPD (B) and CoV (C). Human tetraploid cardiomyocytes showed age-dependent accumulation of sSNVs with an increasing rate of 118 sSNVs per year without correction (A), 109 sSNVs per year with MAPD correction (B), and 94 sSNVs per year with CoV correction(C).



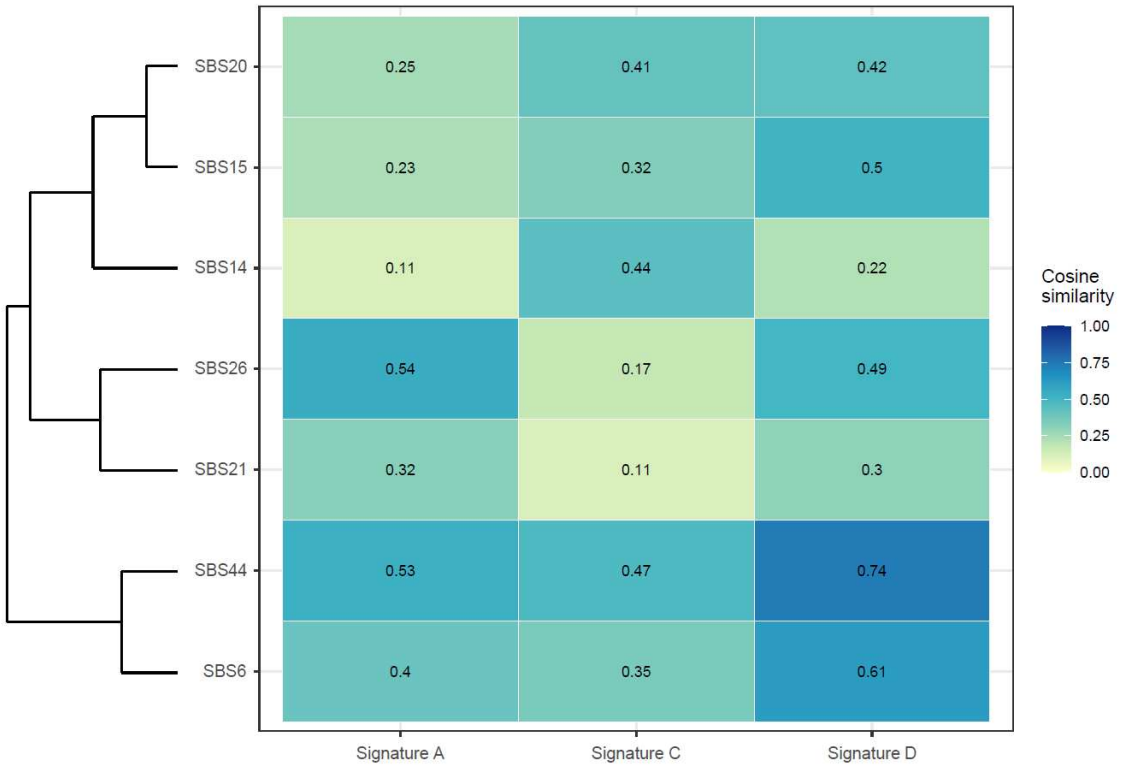
Extended Data Fig. 3. Signature metrics for *de novo* mutational signature analysis. *De novo* mutational signature analysis was performed using non-negative matrix factorization (NMF), in which the factorization rank is critical to define the number of signatures used to decompose the target matrix of sSNVs. We identified that four signatures can maximize the cophenetic and best fit the observed sSNV matrix.



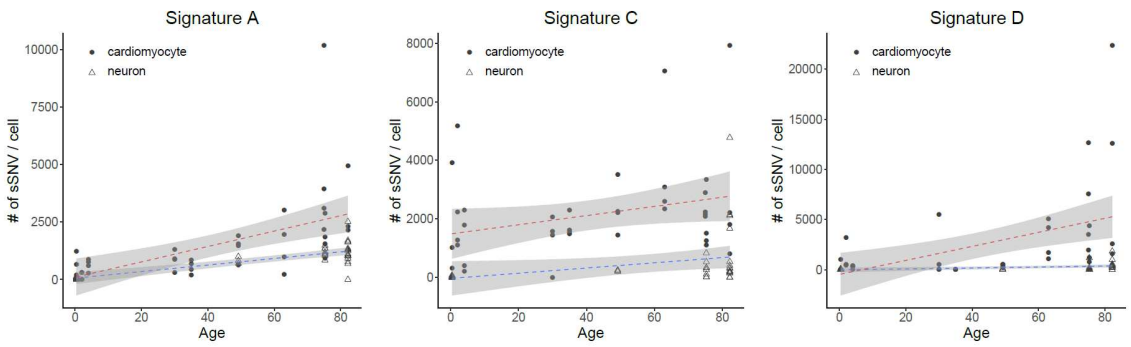
Extended Data Fig. 4. Comparison of mutational signatures identified in this and other studies. (Left panel) *De novo* mutational signatures identified from single human cardiomyocytes in this study. (Middle panel) Previously published signatures identified from single human neurons (Lodato et al.). (Right panel) Recently published signatures thought to represent artefacts of single-cell whole genome amplification, SBS scE and scF, from a study of cultured cells (Petljak et al.). Signature D/N4 was present only in cardiomyocytes but not in neurons. Signature B/N1 identified in cardiomyocytes and neurons resembles the artifact signature SBS scF, thus it was excluded for further mutational burden and signature analyses in this study.



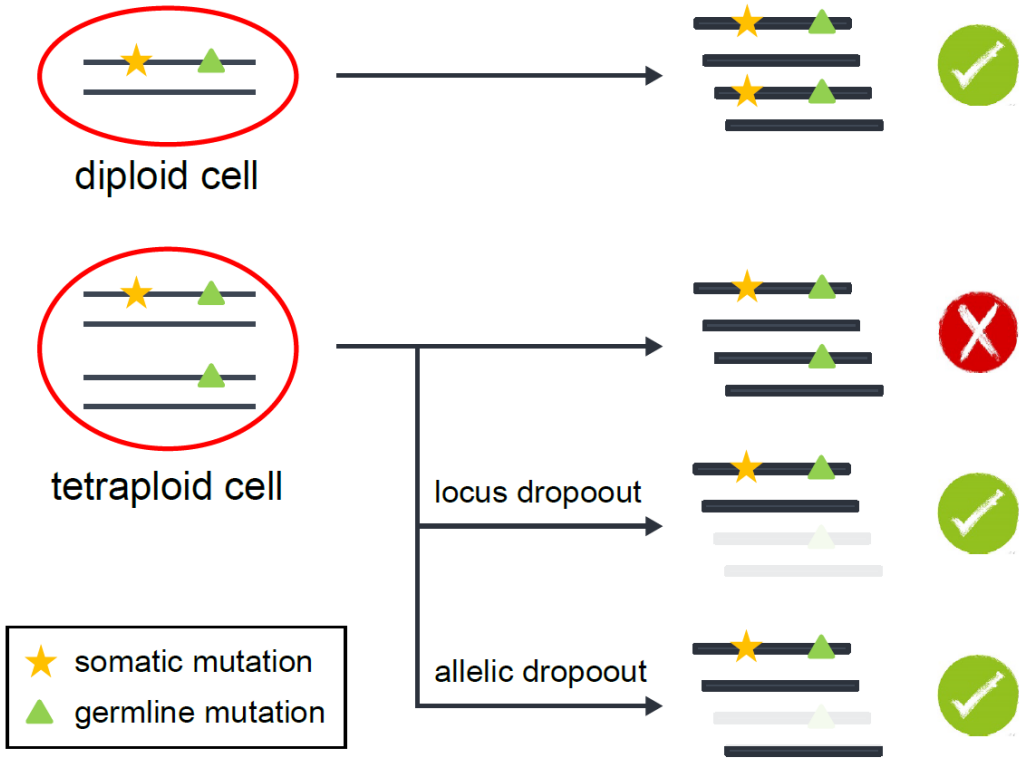
Extended Data Fig. 5. Hierarchical clustering between cardiac and COSMIC cancer signatures. Unsupervised clustering was performed among single-cardiomyocyte-derived signatures (Signatures A, C, and D) and cancer-derived COSMIC signatures (SBS 1-60), by using cosine similarity of 96-class of trinucleotide context to measure the pairwise distance. Signature A resembles to SBS5 and Signature D resembles to SBS44. The red dotted part involving Signatures A, C, and D was highlighted in Figure 3C.



Extended Data Fig. 6. Cosine similarity between single-cardiomyocyte-derived signatures and COSMIC signatures associated with defective mismatch repair (MMR). COSMIC signatures were hierarchically clustered, and cosine similarity values were labeled in each cell with different colors. Signature D showed higher similarity than Signatures A and C in SBS44, SBS6, SBS15, SBS20, and the average across all MMR-associated signatures, indicating its association to defective MMR machinery that fails to fix DNA damages.



Extended Data Fig. 7. Comparative signature analysis between cardiomyocytes and neurons indicating shared and distinct mutational signatures. (Left panel) Signature A related sSNVs showed similar burden between cardiomyocyte and neurons at the birth, but cardiomyocytes accumulated these sSNVs in higher rate with increased age. (Middle panel) Signature C related sSNVs showed generally higher burden in cardiomyocyte but weaker association with age. (Right panel) Signature D related sSNVs burden increased with age distinguishingly only in cardiomyocytes but not in neurons, indicating Signature D is a cardiomyocyte-specific signature.



Extended Data Fig. 8. The model about calling sSNVs from diploid and tetraploid cells by LiRA. In diploid cells, LiRA identified the complete linkage between each sSNV candidate and its adjacent germline heterozygous mutation, which distinguish true sSNVs from technical artifacts. In tetraploid cardiomyocytes, sSNVs present in one out of the four haplotypes were able to be called by LiRA when the reads violating complete linkage were lost due to allelic or locus dropout.

Table S1. Cardiomyocyte with single clade. Number of cardiomyocytes that harbor sSNVs from a single clade among Clades A-I. 25 out of 365 cardiomyocytes that carry sSNVs from two or more clades were shown in Figure 1D-F but not listed here.

Table S2. Ploidy proportion with sensitivity correction. The proportion of single-clade and multi-clade cardiomyocytes before and after sensitivity correction. Odds ratio and P-value (one-tailed proportion test) were calculated for tetraploid and hexaploid+ cells when comparing against diploid cells as negative control.

Table S3. Sample Information and sequencing coverage. 'Case details' tab contains detailed sample information for 10 cases in present study. PMI = Post-Mortem Interval; SIDS = Sudden Infant Death Syndrome; RIN = RNA integrity number. 'Sequencing details' tab contains information about sequencing platform and average coverage.

Table S4. Somatic SNV candidates. Somatic SNVs identified by LiRA from each cardiomyocyte. CC = Composite Coverage, an integer coverage-based quality metric; 'Linked 1K genome SNP' = linked germline heterozygous mutation used for phasing sSNV call; Orientation, whether the somatic alternate allele was on the same haplotype as the germline alternate allele or not.

Table S5. Estimated sSNV counts per cardiomyocyte. Somatic mutation rate per cell estimated by LiRA and adjusted for ploidy status. MAPD (median absolute pairwise difference) and CoV (coefficient of variation) were used for measuring evenness of genome amplification.

Table S6. Exonic sSNVs in cardiomyocytes. Identified 58 exonic mutations with chromosome location and functional annotation.

Figures

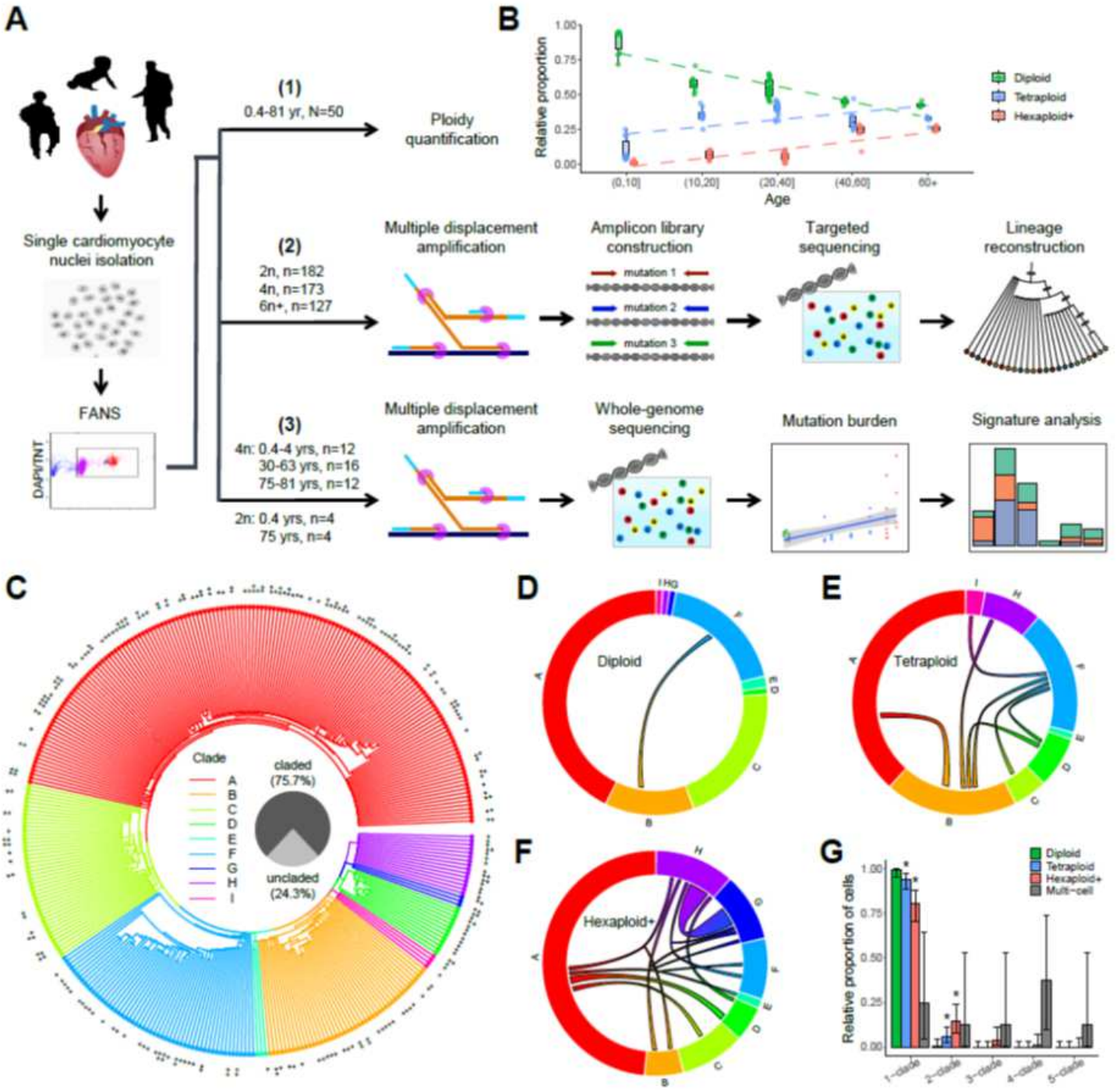


Figure 1

Lineage reconstruction of human cardiomyocytes via single-cell sequencing based on ploidy. (A) Schematic of approach. Nuclei are isolated from frozen postmortem heart tissues with varied ages, and sorted based on ploidy content by FANS. Sorted nuclei are subjected to ploidy quantification (upper), or amplified by Φ29 polymerase-mediated MDA for targeted sequencing for cell lineage analysis (middle) or whole-genome sequencing for sSNV burden and signature analysis (lower) after. The numbers of

examined individuals (N) and nuclei (n) are marked. (B) Increased ploidy in cardiomyocyte with age. Quantification of cardiomyocyte ploidy, from normal postmortem heart of 50 individuals ranging from 0.4 - 82.7 years age, indicating increased ploidy in cardiomyocyte with age. At birth most cardiomyocytes are diploid with only ~5% polyploid cardiomyocytes. The polyploidy proportion increases to ~50% by age 60. Boxplot denotes median with 25% and 75% quartile. (C) Lineage map of 340 single-clade human cardiomyocytes based on panel sequencing of 253 clonal sSNVs. Cardiomyocytes are placed into nine distinct clades defined by one or more clade-specific sSNVs. Tetraploid nuclei and hexaploid or higher (hexaploid+) nuclei are labeled by + and ++, respectively. (D-F) Cardiomyocytes generated by fusion between cells from multiple clades. The connecting arch between clades indicates nuclei containing sSNVs from more than one clade, and the thickness of arch indicates the number of nuclei. All but one diploid nuclei (D) belong to single clade, suggesting the low double-sorting rate, whereas tetraploid (E) and hexaploid+ nuclei (F) have much larger fractions of multi-clade cells generated by fusion. (G) Proportion of cardiomyocytes with different clade numbers. Tetraploid and hexaploid+ cardiomyocytes show higher proportion of multi-clade cells than diploid cells (two-tailed proportion test, asterisk, $p < 0.05$,), suggesting fusion as a mechanism for polyploidization of human cardiomyocytes. Error bar: 95% confidence interval.

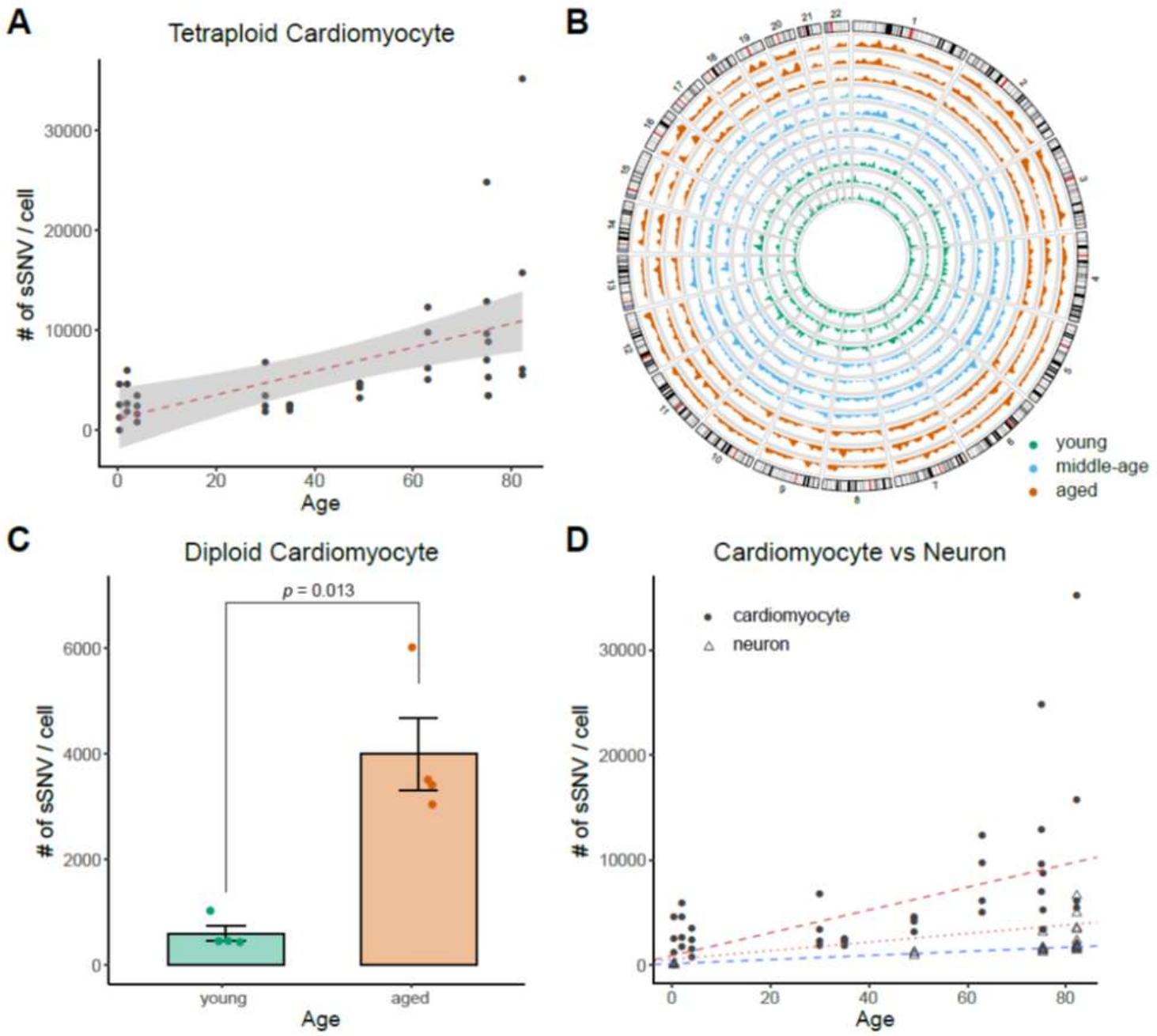


Figure 2

sSNV burden in single human cardiomyocytes detected by whole-genome sequencing. (A) Estimated genome-wide sSNV burden of tetraploid cardiomyocytes. Tetraploid cardiomyocytes showed an increased number of sSNVs with increased age (~2,600 in young vs ~11,500 in aged). Single tetraploid cardiac nuclei were obtained from normal postmortem heart of individuals ranging from 0.4 to 82.7 years of age. (B) Genome-wide distribution of SNVs. Autosomal sSNV density in 10Mb genomic window was plotted for each individual. Green, blue, and red represents young, middle-age, and aged hearts, respectively. (C) Estimated genome-wide sSNV burden of diploid cardiomyocytes. Aged diploid cardiomyocytes showed ~7 fold increase in sSNV burden compared to young cardiomyocytes (~600 in young vs ~4,000 in aged) Error bar: standard error of the mean. Single diploid cardiac nuclei were

obtained from normal postmortem heart of a young (0.4 yrs.) and an aged (75 yrs.) individual. (D) Age-dependent increases of sSNV burden in cardiomyocytes vs neurons. Human cardiomyocytes demonstrate a significantly faster sSNVs accumulation than neurons (~2,600 vs ~100 in young and ~11,500 vs ~2,000 in aged). Red and blue dotted lines indicate the regression lines of cardiomyocytes (thick for tetraploid and thin for diploid) and neurons, respectively.

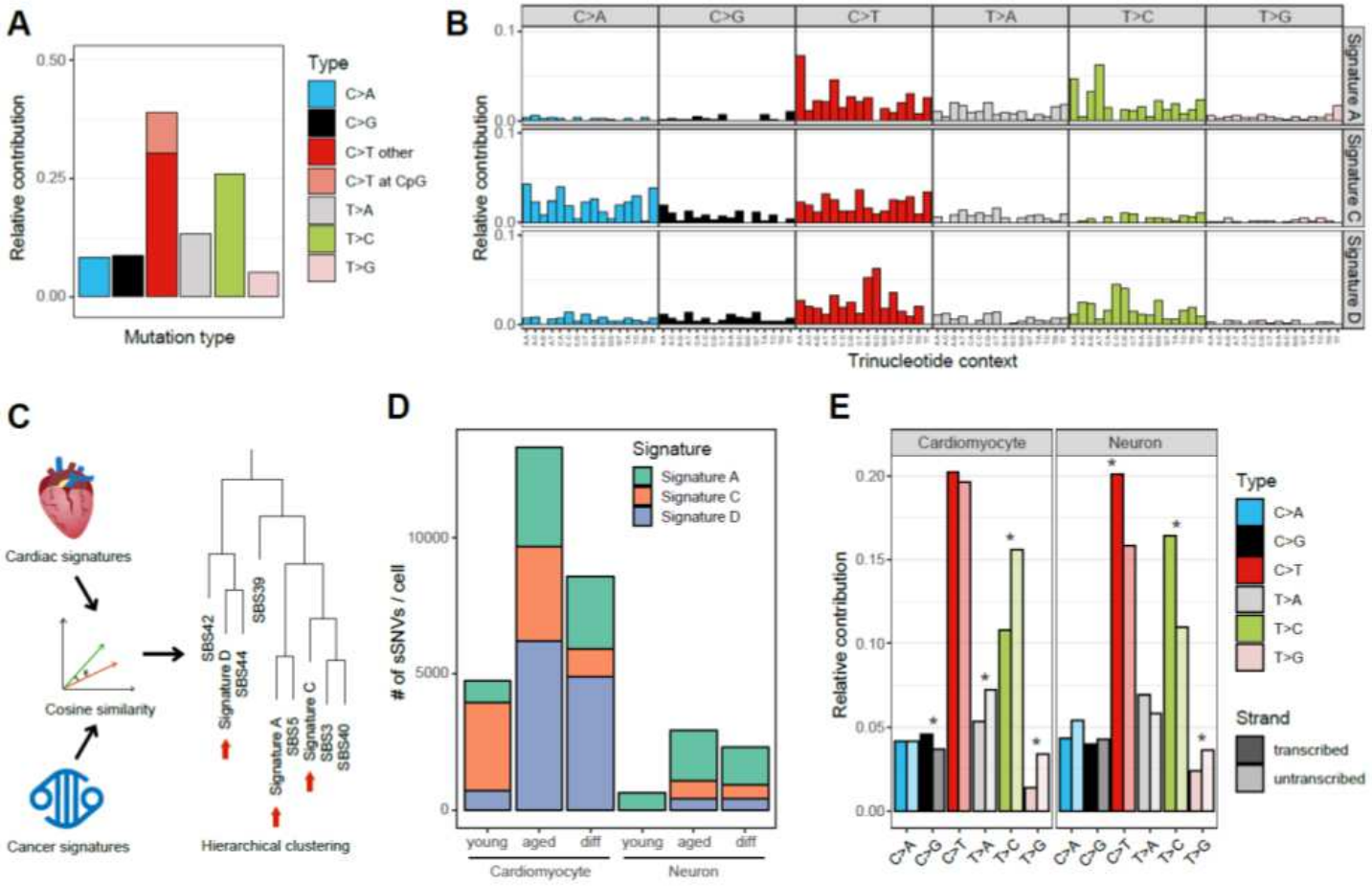


Figure 3

Signature analysis of sSNV reveals mutational process in cardiomyocytes during aging. (A) Substitution type for the age-accumulated sSNVs (the “net increase” of sSNVs between young and aged cardiomyocytes). C>T and T>C substitutions are predominant. (B) Cardiac mutational signatures identified by non-negative matrix factorization based on the trinucleotide context of sSNVs. Each signature is displayed according to the 96 trinucleotide contexts, defined by the six substitution types and sequence context immediately 5’ and 3’ to the mutated base. Although both Signatures A and D predominate with C>T and T>C substitutions, they differ in trinucleotide contexts at C>T and T>C substitutions. (C) Hierarchical clustering between cardiac and COSMIC cancer signatures. Cosine similarity of 96 trinucleotide contexts was used to measure the pairwise distance between signatures. Signature A resembles SBS5 and Signature D resembles SBS44. (D) Signature-specific sSNV burden in human cardiomyocytes and neurons. The “diff” group represents the age-accumulated sSNVs in cardiomyocytes or neurons. Signature D specifically accumulated in aged cardiomyocytes, whereas

Signature A accumulated in both cardiomyocytes and neurons. (E) Transcription bias of age-accumulated sSNVs. Asterisk mark significant difference between transcribed and untranscribed strands ($p < 0.05$, Poisson test). Notably, T>C substitution is enriched in the untranscribed strand in cardiomyocytes whereas it is enriched in the transcribed strand in neurons, suggesting different mutation mechanisms in these two nondividing cells.

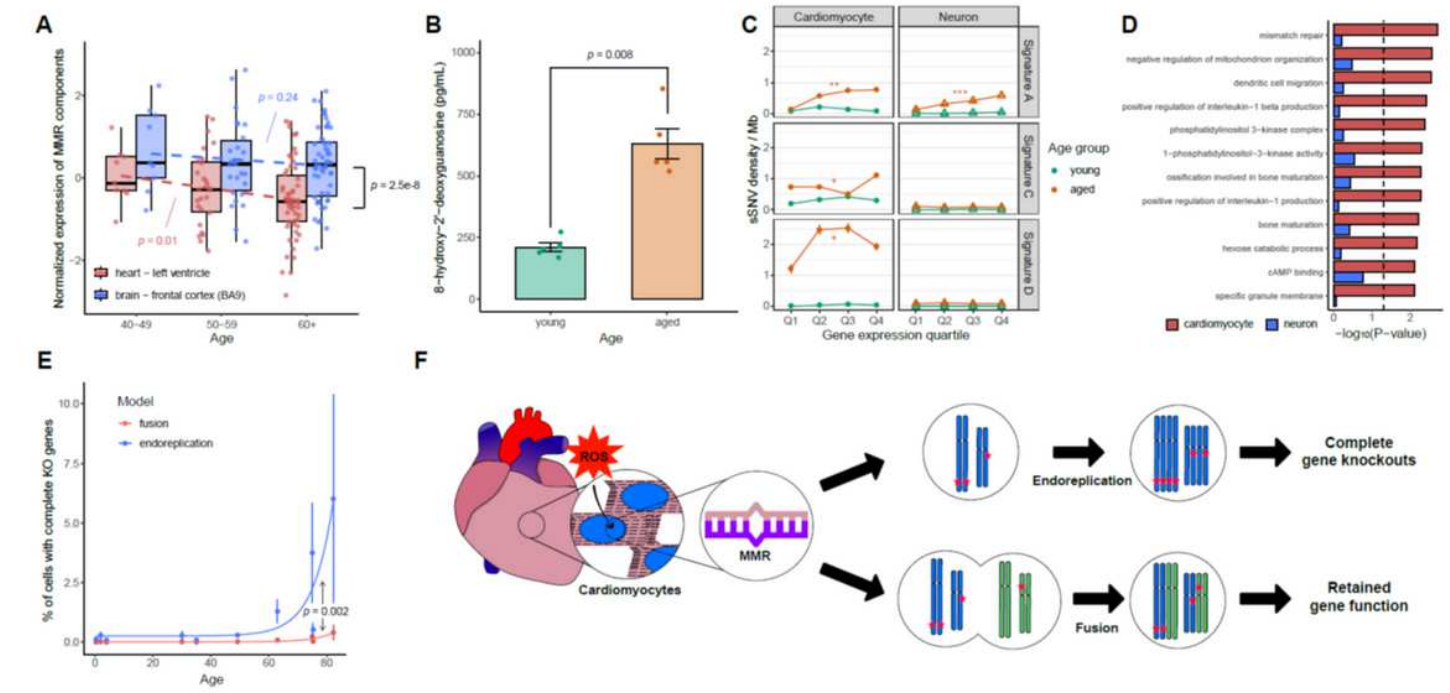


Figure 4

Potential mechanism of sSNV formation and their impact on cardiomyocytes genome during aging. (A) mRNA expression level of MMR complex components in human heart and brain samples of GTEx. Key MMR genes are significantly down-regulated with increased age in heart but not in brain. Boxplot denotes median with 25% and 75% quartiles. (B) Quantification assay of 8-hydroxy-2-deoxyguanosine from human cardiomyocytes. Aged heart shows significantly higher oxidative stress than young heart. Error bar: standard error of the mean. (C) Association of sSNVs with gene expression level. Signature A accumulated more sSNVs in highly expressed genes in both aged cardiomyocytes and neurons. Signatures C and D are dominant in cardiomyocytes without strong correlation with gene expression level. Error bar: 95% confidence interval. +: $p < 0.1$; **: $p < 0.01$; ***: $p < 0.001$. (D) Cardiomyocyte-specifically enriched gene ontology (GO) categories. GO terms with enrichment in cardiomyocytes but not in neurons are listed (FDR-adjusted $p < 0.05$, permutation test). Cardiomyocyte sSNVs are enriched in MMR pathways and pathways involved in metabolism and kinase signaling. (E) Prediction model for effect of sSNVs on abundance of “knockout” tetraploid cardiomyocytes. The percentage of tetraploid cardiomyocytes containing at least one tetra-allelic knock-out gene due to deleterious exonic sSNVs was estimated using fusion and endo-replication models. The fusion model demonstrates a significantly lower “knockout” rate than the endo-replication model, especially in aged heart, suggesting the protective effect of fusion against deleterious sSNV accumulation. Error bar: 95% confidence interval. (F)

Mechanism of sSNV occurrence in heart and adaptation of cardiomyocyte to a fusion mediated polyploid genome. Cardiomyocytes with increased age show increased level oxidative stress (ROS), which results in a mismatched pairing in the genome. Defective MMR in aged cardiomyocytes might not function effectively to repair this increased load of mismatches and lead to fixed sSNVs. Cardiomyocytes with higher ploidy, generated by fusion, can better tolerate the deleterious effect of these mutational burdens.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1ploidywithsingleclade.xlsx
- TableS2polyploidproportion.xlsx
- TableS3caseandsequencingdetails.pdf
- TableS4mutationsiteinfo.xlsx
- TableS5mutationrateinfo.pdf
- TableS6heartexonicmutation.pdf