A model for postzygotic mosaicism quantifies the allele fraction drift, mutation rate, and contribution to de novo mutations

Adam Yongxin Ye,1,2,3,6 Yanmei Dou,1,4,6 Xiaoxy Yang,1 Sheng Wang,4,5 August Yue Huang,1 and Liping Wei1

1Center for Bioinformatics, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, People’s Republic of China; 2Peking-Tsinghua Center for Life Sciences, Beijing 100871, People’s Republic of China; 3Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, People’s Republic of China; 4National Institute of Biological Sciences, Beijing 102206, People’s Republic of China; 5College of Biological Sciences, China Agricultural University, Beijing 100094, People’s Republic of China

The allele fraction (AF) distribution, occurrence rate, and evolutionary contribution of postzygotic single-nucleotide mosaicsisms (pSNMs) remain largely unknown. In this study, we developed a mathematical model to describe the accumulation and AF drift of pSNMs during the development of multicellular organisms. By applying the model, we quantitatively analyzed two large-scale data sets of pSNMs identified from human genomes. We found that the postzygotic mutation rate per cell division during early embryogenesis, especially during the first cell division, was higher than the average mutation rate in either male or female gametes. We estimated that the stochastic cell death rate per cell cleavage during human embryogenesis was ~5%, and parental pSNMs occurring during the first three cell divisions contributed to ~10% of the de novo mutations observed in children. We further demonstrated that the genomic profiles of pSNMs could be used to measure the divergence distance between tissues. Our results highlight the importance of pSNMs in estimating recurrence risk and clarified the quantitative relationship between postzygotic and de novo mutations.

[Supplemental material is available for this article.]

Postzygotic mosaicism describes individuals who developed from a single zygote but consist of multiple cell populations with different genotypes (Strachan and Read 1999). The mosaicism arises due to postzygotic errors in DNA replication (De 2011; Lupski 2013) and can lead to disease states in the mosaic carriers (Poduri et al. 2013; Priest et al. 2016; Terracciano et al. 2016), or in the heterozygous offspring inheriting the mutant allele (Dal et al. 2014; Huang et al. 2014; Acuna-Hidalgo et al. 2015; Dou et al. 2017), and contribute to the recurrent risk of genetic disorders in children (Depienne et al. 2010; Xu et al. 2015; Takahashi et al. 2017). With the recent development of deep sequencing technology, a small number of studies reported the genomic pattern of postzygotic single-nucleotide mosaicsisms (pSNMs) in noncancerous human samples (Huang et al. 2014, 2018; Dou et al. 2017; Ju et al. 2017), and this has enabled the quantitative analysis of pSNMs during the normal developmental process.

Compared to germline mutations, the minor allele fraction (AF) of postzygotic mosaicsisms should deviate from 50%. Since mosaicsisms with varied AFs have different impacts on transmission probability and disease penetrance (Bernkopf et al. 2017; Kono et al. 2017), it is important to quantify the AF distribution of mosaicsisms. Campbell et al. (2014) developed a Galton-Watson process model to predict the recurrence risk of de novo mutation with potential parental mosaicism, by considering both the cell division rate and mutation rate. However, with the limited data at that time, they assumed a constant mutation rate and only quantified the mean and variance of the AF distribution. More recently, Ju et al. (2017) reported the AF distribution of postzygotic mosaicsisms in blood samples from patients with breast cancer and estimated an approximately 2:1 asymmetric contribution of daughter cells during early embryogenesis. However, they assumed the same deterministic bifurcating lineage tree for every embryo, which might be unrealistic due to the occurrence of cell death and variations in cell number in human preimplantation embryos (Hardy et al. 1989, 2001; Mottla et al. 1995). As such, the shape of the AF distribution has not been properly investigated with consideration of the cell death rate during stochastic embryogenesis nor the quantification of the mutation rate of pSNMs and the contribution to de novo mutations.

To address these questions, we developed a new mathematical model to describe the AF distribution. We modeled the dynamics of AF for mosaicsisms by introducing the mutant status of each cell into a previous Galton-Watson branching process accounting for the possibility of cell death. We then generated the theoretical AF distribution by considering the accumulation of mosaicsisms. We fitted our model with two large-scale pSNM data sets from noncancerous individuals (Dou et al. 2017; Huang et al. 2018) to quantitatively characterize the accumulation and AF distribution of pSNMs during normal human development, especially in regard to those arising during early embryogenesis.

© 2018 Ye et al. This article, published in Genome Research, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.230003.117. Freely available online through the Genome Research Open Access option.
Results

Modeling the AF distribution of postzygotic mosaicisms

During the development of the human body, mosaicisms arise due to postzygotic mutations that have escaped from DNA repair machinery. Mutations that occurred in different time stages will lead to different initial AF of mosaicisms. The AF could further drift and deviate from its initial value due to the stochastic process of cell division and death in the development. To quantitatively describe the dynamics of AF drift, we extended the classical Galton-Watson branching process (Hardy et al. 2001) by introducing the initial mutant status of each cell in the initial cell population (Fig. 1A; Supplemental Methods). The branching process assumed cells to have synchronous cleavage, and each cell could either divide with a probability $\alpha$ or die with a probability $\gamma$, or do neither at each cleavage step (Fig. 1A). The behavior of the extended process was determined by the initial cell number ($n_0$) and initial mutant cell proportion ($p_0$), as well as the parameters $\alpha$ and $\gamma$. For a postzygotic mutation, the mutant allele is often present in a fraction of cells as heterozygous genotype and absent in the remaining cells. Therefore, the mutant cell proportion is theoretically equal to $AF \times 2$. We simulated the cleavage branching process and analyzed the distribution of the cell number ($n_i$) and mutant cell proportion ($p_i$) after each cleavage step (Fig. 1B; Supplemental Methods). We found that the mean of cell number after $i$ steps followed the exponential growth $E[n_i] = n_0 \times (1 + \gamma - \alpha^i)$, and the mean of the mutant cell proportion remained the same at the initial mutant cell proportion $E[p_i] = p_0$. The variance of the mutant cell proportion introduced in each step $\Delta \text{Var}[p_i] = \text{Var}[p_i] - \text{Var}[p_{i-1}]$ could be well fitted by the quadratic regression, $C_2 \cdot x_i^2 + C_1 \cdot x_i - \text{where } x_i = p_i \cdot (1-p_i)/n_i$ (Fig. 1C; Supplemental Fig. S1). In addition, the estimated coefficients $C_2$ and $C_1$ could be further fitted as functions of $\alpha$ and $\gamma$ (Fig. 1D, E), as shown in the following equations:

$$C_1(\alpha, \gamma) = 0.2169687 - 0.3578442 \cdot \gamma^2 + 3.6104356 \cdot \alpha^3$$
$$+ 0.5658542 \cdot \gamma^2 - 0.7517724 \cdot \alpha^2 - 0.4307415 \cdot \gamma$$
$$+ 1.9079553 \cdot \alpha - 1.1744103 \cdot \gamma \cdot \alpha (R^2 = 0.999),$$

$$C_2(\alpha, \gamma) = 141.17428 \cdot \alpha^2 - 29.15594 \cdot \alpha^3 + 16.76288 \cdot \alpha$$
$$- 91.07249 \cdot \gamma^2 \cdot \alpha (R^2 = 0.972).$$

With the preceding formulas, we could explicitly calculate the variance of AF drift introduced by each cleavage step for any specific $p$, $n$, $\alpha$, and $\gamma$ and then sum the variances of all the steps to acquire the final cumulative variance. When $\alpha$ and $\gamma$ were constant and $n_i$ was in exponential growth, the first 10 steps contributed to ~90% of the variance (Supplemental Fig. S2; Supplemental Methods), suggesting that most of the variance of the AF drift would come from early cell divisions.

In theory, for the accumulation of mosaicisms when postzygotic mutations were arising, say at the $i$-th cleavage step, the initial AF should be the inverse of the number of haploid genomes ($1/(2 \cdot n_i)$), whereas the number of mosaicisms should be proportional to the number of haploid genomes ($2 \cdot n_i$) and the mutation rate $\mu_i$ (Fig. 1F). After taking the AF drift into consideration, the AF distribution of mosaicisms generated from each cleavage step would expand from the spikes to Gaussian distributions (Fig. 1F). Therefore, we could model the theoretical AF distribution as a Gaussian mixture. Further taking into account the varied detection sensitivities for mosaicisms with different AFs, the observed AF distribution from real sequencing data would be the product of the theoretical distribution and the sensitivity curve (Fig. 1F).

As shown in Supplemental Figure S3, the shape of the AF distribution depended on the relative mutation rate and death rate, suggesting that the AF distribution observed in data might be used to infer these parameters during the cell cleavage process.

We also modeled the amount of de novo mutations transmitted from postzygotic mosaicisms in the parents (transmitted parental mosaicisms). Assuming no proliferative advantage between mutant and wild-type cells, the transmission probability of a mosaicism should be equal to its AF, and the expected number of parental mosaicisms generated in the $i$-th cell division and present in one child was deduced to be equal to the mutation rate $\mu_i$ (per haploid base pair per division) (Supplemental Methods). Compared to the AF distribution of all occurrent mosaicisms, the AF distribution of the transmitted parental mosaicisms would weigh more on mosaicisms with higher AFs, since they would be more likely to transmit the mutant alleles to offspring (Fig. 1G).

Elevated postzygotic mutation rate in early cell divisions, especially in the first division

To estimate the rate of postzygotic mosaicisms in the human genome, we analyzed two large-scale data sets generated by our laboratory. The WGS data set contained whole-genome sequencing of 25 postmortem tissues with no evidence of clonal expansion from five healthy donors, in which 159 nonclonal autosomal pSNMs were identified and validated (Huang et al. 2018). The WES data set contained whole-exome sequencing of 730 families in the Simons Simplex Collection (Fischbach and Lord 2010), in which 187 autosomal pSNMs were identified and validated from 1301 children (Dou et al. 2017). For both data sets, the identified pSNMs were thoroughly validated by targeted ultradepth resequencing (Xu et al. 2015) with an average depth-of-coverage of ~4000-10,000x (Dou et al. 2017; Huang et al. 2018).

We applied a maximum likelihood approach to fit our model using these two data sets (Supplemental Methods). Because evidence shows that characteristic cell death would not appear before the third cell division in human embryos (Hardy et al. 2001), we introduced cell death from the third division and set the death rate $\alpha$ to zero for the first two divisions. We also introduced a “bottleneck” with 50% death rate at the sixth division (about 16 cells randomly selected from 32 cells, and then dividing to 32 cells if in complete bifurcation), according to the differentiation of inner cell mass during the blastocyst stage (Hardy et al. 1989). In addition, because it was reported that the mutation rate might be higher during the first cell division in Drosophila (Gao et al. 2011), we allowed the mutation rate of the first cell division to be different from the mutation rate of the latter divisions and set $\kappa$ as the relative ratio between the first and latter divisions. As shown in Figure 2A and B, the maximum likelihood estimate (MLE) of death rate $\alpha$ was ~0.04-0.06 in both data sets, and $\kappa$ was estimated to be ~2-2.4, with the 10% likelihood interval not overlapping $\kappa = 1$. Consistent with the observation in Drosophila (Gao et al. 2011), our analysis suggested a significantly elevated mutation rate of the first cell division relative to the latter divisions in human embryos. We then merged both data sets and found the MLE of $\alpha$ and $\kappa$ was 0.05 and 2.15, respectively, which fitted both data sets well; these were used in further analyses (Fig. 2C,D). With the estimated parameters (Supplemental Table S1) and correction for the pSNM detection sensitivity (Supplemental Fig. S4), we estimated the mutation rate to be ~8 x 10^{-10} per haploid base pair per division for the first cell division and ~3.7 x 10^{-10} for later divisions.
Figure 1. Model describing the accumulation and allele fraction (AF) of postzygotic mosaicisms. (A) The extended Galton-Watson branching process for cell cleavage and AF drift. In each synchronized cleavage step, a cell could die with a probability $\alpha$ or divide with a probability $\gamma$. Mutant status (gray) is introduced for mosaicisms, summarized as mutation cell proportion ($p$) with cell number ($n$) as parameters. (B) The simulated joint distribution of cell number ($n_i$) and mutant cell proportion ($p_i$) after the $i$ cleavage steps for each combination of initial parameters ($\alpha$, $\gamma$, $n_0$, and $p_0$). (C) The quadratic regression of the increment of variance of mutant cell proportion $\Delta \text{Var}[p] = \text{Var}[p_i] - \text{Var}[p_{i-1}] = C_2 \cdot n_i \cdot x_i + C_1 \cdot x_{i-1}$, where $x_i = p_i \cdot (1-p_i)/n_i$. (D) The regression of the fitted coefficients $C_1$ and $C_2$ on the combination of $\alpha$ and $\gamma$. The colored circles are the sample points, and the black dots show the fitted values. Different colors indicate different $\gamma$ in the plot with $\alpha$ as the x-axis and different $\alpha$ in the plot with $\gamma$ as the x-axis. (E) The formulas $C_1(\alpha, \gamma)$ and $C_2(\alpha, \gamma)$ predict $C_1$ and $C_2$ well. The blue line is the diagonal line. (F) The expected positions of the initial AF and the relative amount for mosaicisms generated in each cleavage step, assuming a constant mutation rate for demonstration. (G) Theoretically, the relative ratio for naturally occurring mosaicisms should be proportional to the number of haploid genomes (similar to “exponential growth”). (G) If we consider parental mosaicisms present in one child as de novo mutations, the relative ratio for each cleavage step should stay as 1. The inner bell-shape curves with different colors show the components for different cleavage steps. Parameters were set as $\alpha = 0.05$ and $\gamma = 0.95$, assuming no bottleneck and a constant mutation rate, for demonstration.
Maximum likelihood estimation (MLE) of cell death rate ($\alpha$) and relative mutation rate ratio ($\kappa$) with the observed AF distribution. (A,B) Contour plots of the likelihood of our model fitted on the (A) WES and (B) WGS data sets. The x-axis denotes the death rate ($\alpha$), and the y-axis denotes the relative ratio of the mutation rate between first division and latter divisions ($\kappa$). The MLEs of $\alpha$ and $\kappa$ and the corresponding log likelihood are labeled with a cross. The curves from inside to outside show 10%, 1%, and 0.1% likelihood intervals, respectively. Aside from $\alpha$ and $\kappa$, the division rate ($\gamma$) is free to change, whereas the other parameters “mut steps,” “death from,” “bottleneck at,” bottleneck $\alpha$, and bottleneck $\kappa$ are set to 7, 3, 6, 0.5, and 0.5, respectively (Supplemental Table S1; Supplemental Methods). (C,D) Histogram of the AF distributions observed in the (C) WES and (D) WGS data sets. The thick brown curves denote the MLE-fitted, observed AF distribution with $\alpha = 0.05$, $\gamma = 0.95$, and $\kappa = 2.15$. The thin inner bell-shape curves with different colors denote the components for different cleavage steps.

Figure 2. Maximum likelihood estimation (MLE) of cell death rate ($\alpha$) and relative mutation rate ratio ($\kappa$) with the observed AF distribution. (A,B) Contour plots of the likelihood of our model fitted on the (A) WES and (B) WGS data sets. The x-axis denotes the death rate ($\alpha$), and the y-axis denotes the relative ratio of the mutation rate between first division and latter divisions ($\kappa$). The MLEs of $\alpha$ and $\kappa$ and the corresponding log likelihood are labeled with a cross. The curves from inside to outside show 10%, 1%, and 0.1% likelihood intervals, respectively. Aside from $\alpha$ and $\kappa$, the division rate ($\gamma$) is free to change, whereas the other parameters “mut steps,” “death from,” “bottleneck at,” bottleneck $\alpha$, and bottleneck $\kappa$ are set to 7, 3, 6, 0.5, and 0.5, respectively (Supplemental Table S1; Supplemental Methods). (C,D) Histogram of the AF distributions observed in the (C) WES and (D) WGS data sets. The thick brown curves denote the MLE-fitted, observed AF distribution with $\alpha = 0.05$, $\gamma = 0.95$, and $\kappa = 2.15$. The thin inner bell-shape curves with different colors denote the components for different cleavage steps.

Parental mosaicsisms occurring during the first three cell divisions contribute to ~10% of the de novo mutations in children

In clinical applications, a heterozygous mutation is defined as “de novo” if the mutant allele is undetectable in both parents. Previous studies have demonstrated that some mutations thought to be “de novo” were actually inherited from parental mosaicsisms that were missed by conventional Sanger sequencing (Jones et al. 2001; Depienne et al. 2006; Chen et al. 2014). Based on our model, assuming that most mutations are neutral and the mutation rate is constant for each division, the contribution of parental mosaicsisms to de novo mutations would be linearly proportional to the number of cell divisions in which the detectable mosaicsisms occurred, and parental mosaicsisms occurring in the first 10 cell divisions would contribute to ~5%-10% of de novo mutations. (Supplemental Fig. S5). With the elevated mutation rate for early cell divisions estimated from the pSNM data sets, the contribution of early parental mosaicsisms would be increased, which would reach to ~10% for parental mosaicsisms that had occurred in just the first three cell divisions (with AF greater than ~4%) (Supplemental Fig. S5).

Theoretically, parents with detectable mosaicism should have a much higher possibility to have children with recurrent de novo mutations. Using a Bayesian framework based on our model, we estimated the risk of recurrence for de novo mutations that were already observed in one child. Similar to previous estimates (Campbell et al. 2014), when assuming a constant mutation rate for each division, the risk of recurrence for de novo mutations already observed in one child would be 1.7% for maternal origin.
and 0.13% for paternal origin (Table 1). However, when we considered the elevated mutation rate in early cell divisions, the risk of recurrence would be as high as 4.8%–11.5% for maternal origin and 1.3%–1.6% for paternal origin (Table 1), which is an approximately threefold to 10-fold elevation compared to that of a constant mutation rate. Our model also demonstrated the value of screening for parental mosaicisms during genetic counseling (1) if a corresponding parental mosaicism was detected with confirmed AF \( \theta_{\text{exact}} \), the recurrence risk would equal to \( \theta_{\text{exact}} \) otherwise, (2) if a corresponding parental mosaicism was not detected by a detection method with a lower AF threshold \( \theta_i \), the recurrence risk would be reduced to approximately \( \theta_i/2.5 \), and this would be higher for mosaicisms of maternal origin compared to that of paternal origin (Supplemental Fig. S6). For instance, considering that conventional Sanger sequencing can only detect mosaicisms with an AF of 5% or more (Depienne et al. 2006; Chen et al. 2014; Xu et al. 2015), screening parental mosaicism by Sanger sequencing might reduce the risk of recurrence to \( \sim 0.67\%–1.8\% \) for maternal origin or \( \sim 0.15\%–0.18\% \) for paternal origin.

Distance measure between tissues based on the AF of shared mosaicisms

After a postzygotic mutation has occurred, all daughter cells would carry the mutant allele. Therefore, the similarity of profiles for postzygotic mosaicisms could theoretically reflect the cell lineage tree across different tissues from the same individual. Our model predicted that mosaicisms with different AFs should have different variances for AF drift; thus, the Euclidean distance between vectors of AFs for shared mosaicisms is theoretically improper due to the different variance between elements. To stabilize the variance of AFs, we proposed Var[\( (p_2 - p_1)/\sqrt{p_0(1-p_0)} \)] as the measure of inter-tissue distance, where \( p_0, p_1, \) and \( p_2 \) are the mutant cell proportion (AF \( \times 2 \)) in the ancestor cell population, one tissue sample, and the other tissue sample, respectively (Methods; Fig. 3A). This distance was solely based on the AF of shared pSNMs between tissues, assuming that each tissue had undergone an independent developmental process from the most recently common ancestral cell population (Fig. 3A).

To assess the performance of this pSNM-based measure, we applied it against the pSNM profiles of five individuals from the WGS data set. Considering that the developmental process was generally identical across individuals, our method enabled us to combine the pSNM lists identified from the specific tissues of multiple individuals. We clustered the tissues from males and females separately based on the pairwise distance matrix between tissues. As shown in Figure 3, B and C, colon and liver were closely clustered, whereas skin shared more similarity with prostate or breast than brain, colon, or liver. These results were in accordance with the knowledge that colon and liver are endodermal, brain is ectodermal, whereas skin, breast, and prostate are mixtures of ectodermal and mesodermal origins (Pansky 1982; Argani et al. 1998; Gilbert 2003).

Table 1. Risk of recurrence of a de novo mutation already observed in one child

| Mutation rate and whether screening for parental mosaicisms or not | Paternal origin | Maternal origin | Unknown origin (~75%–90% paternal) |
|---|---|---|---|
| Screened and detected parental mosaicism with AF = \( \theta_{\text{exact}} \) | Both parents not screened | 0.13% | 1.7% | 0.54%–0.29% |
| Constant mutation rate: \( \mu_i(1) = \mu_i(2) = \ldots = \mu_i(\gamma_{\text{parental}}) = \mu_i(d)/\gamma_{\text{parental}} \) | Both parents screened and neither detected by Sanger sequencing | 0.037% | 0.53% | 0.16%–0.087% |
| Elevated early mutation rate: \( \mu_i(1) = 8 \times 10^{-15}, \mu_i(2) = \mu_i(3) = \ldots = 3.7 \times 10^{-15} > \mu_i(d)/\gamma_{\text{parental}} \) | Both parents not screened | 1.6%–1.3% | 4.8%–11.5% | 2.4%–2.3% |
| | Both parents screened and neither detected by Sanger sequencing | 0.18%–0.15% | 0.67%–1.8% | 0.30%–0.31% |

Figure 3. Measuring inter-tissue distance based on the AF similarity of shared mosaicism. (A) The AF difference of shared mosaicisms in two tissues could be traced back to their most recent common ancestral cell population, but not to an earlier stage when the mosaicisms were generated. (B, C) Clustering trees based on the pairwise distance matrix estimated from the WGS data set for (B) males and (C) females. The bootstrapping values labeled on the internal branches show the bootstrap supporting percentage on the partition of that branch.
in a growing population. Previous studies in population genetics had developed the Wright-Fisher model (Fisher 1922; Wright 1931, 1939) or Moran model (Moran 1958) to describe the genetic drift, and diffusion approximation can be applied in a large population (Kimura 1954, 1955). However, a human embryo develops from a zygote and contains only a few cells in the first several divisions of embryogenesis, and the approximate linear formula might be unsuitable for such a small population size (Supplemental Fig. S1). Here, we proposed a quantitative model for describing postzygotic mosaicisms, which was derived from the Galton-Watson branching process that had been commonly used to model the dynamics of rare alleles and exponential population growth. Since our model focused on neutral postzygotic mosaicisms, we applied our model to analyze the AF distribution of pSNMs from two data sets after excluding those sites with evidence of clonal expansion (Dou et al. 2017; Huang et al. 2018), because they might be driven by selective advantage between mutant and wild-type cells (Abyzov et al. 2017).

Recently, Ju et al. (2017) reported a decrease of 25% AF on the AF distribution and thus explained the phenomenon by the asymmetric contribution of daughter cells resulting from a bottleneck during early embryogenesis. However, we did not observe a decrease of 25% AF in the two data sets we used (Fig. 2C,D). Based on our simulation, a decrease at expected initial AF could only be led by asymmetric contribution restricted to a small range deviated from 0.5, or a nonrandom bottleneck in complete bifurcating cleavage (Supplemental Fig. S7). When introducing deterministic asymmetric contribution into our model, the estimated contribution of the two daughter cells at the two-cell stage was about 0.60.4. However, the fitting was usually not significantly better than the symmetric setting (likelihood ratio test), except for the most flexible model (Supplemental Table SS; Supplemental Methods). In addition, the sensitivity for detecting mosaicism in Ju et al. (2017) was lower (peak value <5%) than our method (Huang et al. 2017). Further studies with a larger sample size and better detection methods might be required to resolve these conflicting observations.

We estimated the embryonic postzygotic mutation rate as being $8 \times 10^{-10}$ per haploid base pair per division for the first cell division and $3.7 \times 10^{-10}$ for latter divisions, whereas the average mutation rate estimated from de novo mutations was $4.2-5.1 \times 10^{-11}$ per haploid base pair per division in germline cells from males versus $0.8-2.4 \times 10^{-10}$ in germline cells from females. The observed elevated mutation rate during the early cell divisions and the first cell division was robust when assuming complete bifurcating cleavage and allocation of pSNMs into different AF groups (Supplemental Table S6). In addition, we examined and found that adding noise to observed AF, relaxing some parameter constraints in maximum likelihood estimation, allowing changeable death rate and division rate, or taking asymmetric contribution into consideration did not affect the two main findings: (1) Postzygotic mutation rate during early cell divisions was higher than the average mutation rate per cell division estimated from de novo mutations; and (2) the mutation rate of the first division was even higher than the mutation rate of the latter divisions (Supplemental Table SS; Supplemental Fig. S8; Supplemental Methods). Furthermore, we found that introducing changeable death rate and division rate did not fit the AF distribution significantly better ($P$-value $>0.6$, likelihood ratio test).

Our model demonstrated that postzygotic mosaicisms could be regarded as a partition along the mutation process in the germ-line lineage, and the amount of mosaicisms that can be detected should exponentially increase with a lower detection threshold of AF. We reported that early parental mosaicisms in the first three cell divisions contributed to ~10% of de novo mutations in the offspring, which was close to a previous estimation of 8.6% observed for Dravet syndrome (Xu et al. 2015). We further highlighted an elevated risk of recurrence for de novo mutations in the offspring when early divisions had elevated mutation rates. Since mutations with different AFs showed varied impacts on the disease phenotype (Meng et al. 2015), the AF distribution of mosaicisms could be different between healthy and disease-associated genomes. Theoretically, we could extend our model by introducing more parameters, including prebirth lethality $1-\nu(\theta)$, post-birth penetrance $\pi(\theta)$, and fertility $r(\theta)$ (Supplemental Table S7; Supplemental Methods), but the accurate estimation of such parameters requires more data from different disease states.

For screening parental mosaicism, it was ideal to use gametes, such as sperm samples, rather than blood samples. As shown in previous findings, mosaicisms with high AF were probably shared in multiple tissues, indicating their early arising during embryogenesis and the usage of blood sample for screening parental mosaicisms (Huang et al. 2014; Yang et al. 2017). Since blood stem cells and germline stem cells are known to segregate at roughly the 15th division (Campbell et al. 2014), germline mosaicisms with low AF that have occurred after blood-germline segregation (AF lower than $\sim3 \times 10^{-5}$) could be only detectable in gametes but not blood samples. In addition, mosaicisms may occur later and rise to high AF as a result of clonal expansion (Goriely et al. 2013; Xie et al. 2014; Huang et al. 2018), which has not been modeled in this work. Screening parental mosaicisms from blood samples may miss the germline mosaicisms under selfish selection, which may cause serious diseases such as Costello syndrome (Goriely et al. 2013). Therefore, the AF of mosaicisms in germline and blood samples may be different, and the results of clinical application of blood test for parental mosaicism should be interpreted with caution.

We noticed that our distance measure based on AF of shared mosaicisms was similar to $F_{ST}$ and the $F$-statistics applied in population genetics (Reich et al. 2010; Patterson et al. 2012). All of these metrics used $p \cdot (1-p)$ as the denominator to normalize the metric, which was required when attempting to combine information from different sites with different frequencies. The distance measure could also be calculated by the procedure proposed for normalized $F$-statistic (Reich et al. 2010; Patterson et al. 2012), which gave different weight for each site and produced results similar to our procedure (Supplemental Table S9; Supplemental Methods). We inferred the lineage tree of multiple types of tissue obtained from five donors, assuming the developmental process of tissues is generally identical across individuals. Our result suggested that skin was clustered with prostate or breast rather than brain. Considering that our skin samples consisted of more dermis than epidermis, this clustering pattern could be explained by the shared mesodermal origin between dermis, prostate, and breast. $F$-statistics have been demonstrated to infer mixture in population genetics, which could be potentially applied to study mosaicisms.

In this study, we modeled the accumulation and AF distribution of postzygotic mosaicisms by formulating the AF drift and generating a theoretical AF distribution as a Gaussian mixture, with consideration of cell death during zygote cleavage. Our model provided an estimation regarding the occurrence rate of postzygotic mutations and highlighted their roles in the origin of de novo mutations. With the development of next-generation sequencing techniques and growing data sets regarding postzygotic mosaicisms, we should be better able to test our model and
describe the dynamics of postzygotic mosaicsisms in the future. Our work sheds new light on the quantitative characterization of postzygotic mosaicsisms in human development and provides guidance for screening mosaicism for use in clinical applications.

Methods

Estimation of mosaicism occurrence rate

To estimate the mutation rate, we calibrated the psNM detection sensitivity for the WGS and WES data sets (Supplemental Fig. S4) by mixing, in silico, the sequencing reads of two well-genotyped individuals, NA12878 and NA12891 (Supplemental Table S8), following published protocols (Huang et al. 2014, 2017). We quantified the number of effective base pairs after filtering out repeat regions. Considering the relatively low sensitivity for detecting psNMs from low-coverage regions (Huang et al. 2014, 2017), we only focused on the sites with depth \( \geq 40 \times \) (baseQ \( \geq 20 \) and mapQ \( \geq 20 \)). This filter was also applied to the validated mosaicsisms in the WGS data set (Huang et al. 2018) and the WES data set (Dou et al. 2017).

To compare the mosaicism occurrence rate and the de novo mutation rate, we unified the unit to the mutation rate per haploid base pair per cell division. The average mutation rate per haploid base pair per division was calculated by dividing the de novo mutation rate per (paternal or maternal) haploid base pair per generation \( \mu_g \) by the cell division number from zygote to sperm in the father or from zygote to egg in the mother (Supplemental Table S4) as follows:

\[
\bar{\mu}_g = \frac{\mu_g}{\text{egg}\_\text{gamete}} - \frac{\mu_g\_\text{paternal}}{\text{paternal}\_\text{gamete}} - \frac{\mu_g\_\text{maternal}}{\text{maternal}\_\text{gamete}}, \quad \text{for father}
\]

\[
\bar{\mu}_g = \frac{\mu_g}{\text{egg}\_\text{gamete}} - \frac{\mu_g\_\text{paternal}}{\text{paternal}\_\text{gamete}} - \frac{\mu_g\_\text{maternal}}{\text{maternal}\_\text{gamete}}, \quad \text{for mother}
\]

The data set we used for estimating average mutation rate is the same WES data set from Simons Simplex Collection (Fischbach and Lord 2010), with detailed information of parental ages at children’s birth (Iossifov et al. 2014, Supplementary Table 1). For the cell division number in the father, we took the paternal age into consideration and estimated the cell division number for each father with the formula \( 34 + 23 \times \text{father} \_\text{age at child birth} - 38 \times 7/365 - 15 \) + 4 (Rahbari et al. 2016, Supplementary Figure 1).

For psNMs, the detection sensitivity varied across the different AF (Huang et al. 2014, 2017), which should be taken into consideration. We introduced constant death rate (a) and division rate (\( \beta \)) from the third division, leaving the first two divisions completely bifurcating (set \( a = 0 \) and \( \beta = 1 \)). We introduced the relative ratio of the mutation rate for the first cell division to the mutation rate of latter divisions (\( \delta \)). We estimated \( a, \beta, \delta, \) and \( \gamma \) from the two data sets using a maximum likelihood approach (Supplemental Methods). We then estimated the mutation rate by dividing the amount of psNMs allocated for each cell division by the effective base pair number and expected number of haploid genomes, after calculating the theoretical proportion of psNMs for each cell division adjusted for the sensitivity and the probability reaching fixed or lost (AF reaching 0.5 or 0) (Supplemental Table S2). For details, see Supplemental Methods.

Contribution of parental mosaicsisms to de novo mutations in children

Assuming no selection between mutant and wild-type cells, our model showed that the expected amount of parental mosaicsisms occurred at the \( i \)-th cleavage step and transmitted to one child was equal to \( \mu_g^{(i)} \), the mutation rate per haploid base pair for the \( i \)-th cell division. Denote \( \mu_d \) as the de novo mutation rate per haploid base pair per generation, then \( \mu_d = \sum_{i=1}^{\text{egg}} \mu_g^{(i)} \), where \( \text{egg}_\text{gamete} \) is the number of cell divisions along the germline lineage from zygote to gamete. Therefore, the contribution of parental mosaicsisms occurring in the \( i \_d \sim i \_k \)-th cell division to de novo mutations in children could be written as \( \sum_{i_d}^{\text{egg}} \mu_g^{(i)} / \sum_{i_1}^{\text{egg}} \mu_g^{(i)} \), which could be regarded as a partition along the mutation process in the germline lineage. With the estimated elevated mutation rate for psNMs \( \mu_g(1) = 8 \times 10^{-10} \) for the first cell division, and \( \mu_g(2) \approx \mu_g(3) \approx \mu_g(4) \approx 3.7 \times 10^{-10} \) for later divisions, after calculating \( \mu_g^{(i)} \) adjusted for fixation or loss (Supplemental Table S2), the proportion of two times (for both parents) the cumulative mutation rate in the diploid de novo mutation rate \( 2\mu_g(1) + \mu_g(2) + \ldots/l/2\mu_g \) was calculated, as shown in Supplemental Figure S5. For details, see Supplemental Methods.

Recurrent risk

The risk of recurrence for de novo mutations could be calculated in a Bayesian framework derived from the mosaicism accumulating process. The recurrence risk of a mosaicism already known to have been transmitted to one child can be deduced as

\[
RR^{(1)} = \int \frac{1}{2m(\theta)} \sum_{i=1}^{\text{egg}} \frac{\mu_g^{(i)}}{\mu_g} \frac{1}{2m(\theta)} = \sum_{i=1}^{\text{egg}} \frac{\mu_g^{(i)}}{\mu_g} \frac{1}{2m(\theta)}
\]

When the mutation has been screened to confirm the mosaic AF status in both parents,

1. If we detect the mosaicism with AF = \( \theta \text{exact} \) in the gametes of either parent, then the recurrence risk will become \( \theta \text{exact} \).

2. If we could not detect the mosaicism with AF \( \geq \theta_i \) in either parent, then the recurrence risk will become

\[
RR^{(2)} = \sum_{i=1}^{\text{egg}} \frac{\mu_g^{(i)}}{\mu_g} \frac{1}{2m(\theta)}
\]

The paternal contribution to the origin of a de novo mutation was set to \( 75\% - 90\% \) (Venn et al. 2014; Yuen et al. 2016), which influenced the estimate of the de novo mutation rate \( \mu_d \) per paternal or maternal haploid. For numerical estimation of recurrence risk, here we set the number of cell divisions \( \text{egg}_\text{gamete} = 30 \) for maternal egg or 400 for paternal sperm, which was estimated assuming 30-year-old parents (Rahbari et al. 2016). For a constant mutation rate, the mutation rate \( \mu_g^{(i)} \) was set to the average mutation rate \( \mu_d/\text{egg}_\text{gamete} \), assuming no bottleneck. For an elevated early mutation rate, the risk of recurrence was conservatively estimated assuming that the mutation rate of all early divisions was \( 3.7 \times 10^{-10} \) (per haploid base pair per division), with the exception of \( 8 \times 10^{-10} \) for the first division, which might coerce the later mutation rate close to 0. For Sanger screening, the lower detection limit was set to \( \sim 5\% \); therefore, we set the corresponding \( \theta_i \approx 3 \). The numerical results are shown in Table 1. For details, see Supplemental Methods.

Measuring the distance between tissues by the variance of the AF difference between shared mosaicsisms

When the initial cell number of a tissue differentiated from the common ancestor cell population was large (\( n_0 \geq 40 \)), the quadratic term of AF drift \( C_2 \cdot \Delta x^2 \) could be omitted, and the final AF of a neutral mosaicism in a tissue (\( p_1/2 \)) would approximately follow the normal distribution (Fig. 3A)

\[
p_1 \sim N(p_0, \frac{p_0(1-p_0)}{n_0}, \frac{C_2(\alpha_1, \gamma_1)}{(1 + \gamma_1 - \alpha_1)}^	op)
\]

Genome Research 949

www.genome.org
The expectation stays at $p_0$ and the variance can be decomposed into two parts: $p_0 - (1-p_0)$, which is dependent on the AF itself, and

$$d_l := \frac{1}{n_{d,1}} \sum C_i (n_i^1, n_i^2) (1 + n_i^1 - n_i^2)^{-1}$$

which represents the developmental process from the ancestor cell population to the mature tissue. Assuming the two tissues developed independently after differentiation from their most recent common ancestral cell population, we would have

$$(p_2 - p_1) \sim N(0, p_0 (1 - p_0) \cdot (d_1 + d_2))$$

$$\Rightarrow \text{Var}[p_2 - p_1] = E[(p_2 - p_1)^2] = p_0 (1 - p_0) \cdot (d_1 + d_2)$$

Since the distance $(d_1 + d_2)$ is dependent only on the developing processes of the two tissues and should be stable for a specific pair of tissues, we proposed a distance measure

$$\text{Var} \left[ \frac{p_2 - p_1}{\sqrt{p_0 (1 - p_0)}} \right] = E \left[ \left( \frac{p_2 - p_1}{\sqrt{p_0 (1 - p_0)}} \right)^2 \right]$$

to estimate $(d_1 + d_2)$ between tissues, which could normalize and summarize the AF difference of each neutral mosaicism shared in two specific tissues to estimate the relative distance between tissues.

In the real WGS data, we used $p_0 = (p_1 + p_2)/2$ as a simple guess of $p_0$, and $\text{Var} \left[ \frac{p_2 - p_1}{\sqrt{p_0 (1 - p_0)}} \right]$ was calculated using

$$\text{MAD} \left[ \frac{p_2 - p_1}{\sqrt{p_0 (1 - p_0)}} \right]^2 \times \text{median}(\text{abs}(x))$$

where median absolute deviation $\text{MAD}(x) = 1.4826 \times \text{median}(\text{abs}(x))$, to make the estimate less sensitive to outliers.

Based on the pairwise distance matrix between multiple tissues, hierarchical clustering was carried out using the Ward agglomeration method. To evaluate the confidence of the clustering results, we performed bootstrapping by resampling the mosaics shared between two tissues and obtained a bootstrap distribution of the distance. We reported the clustering trees in which all the bootstrapping values were 65 or greater. For details, see Supplemental Methods.

Acknowledgments

We thank Dr. Jian Lu for his valuable comments and suggestions. We thank Drs. Jianui Li, QiXi Wu, Stephen J. Sanders, and Matthew W. State for their enlightening discussion and comments. We also appreciate the reviewers for their helpful suggestion and comments. This work was supported by the Ministry of Science and Technology of the People's Republic of China (2015AA02010108) and National Natural Science Foundation of China (Grant No. 31530092).

References

Abyzov A, Tomasinli L, Zhou B, Vasmatzis N, Coppola G, Amendini M, Pattini R, Wilson M, Gerstein M, Weissman S, et al. 2017. One thousand somatic SNVs per skin fibroblast cell set baseline of mosaic mutational load with patterns that suggest proliferative origin. Genome Res 27: 512–523.

Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, Hotschen A, Vissers LE, Gilissen C. 2015. Post-zygotic point mutations are an underrecognized source of de novo genomic variation. Am J Hum Genet 97: 67–74.

Argani P, Walsh PC, Epstein JI. 1998. Analysis of the prostatic central zone in patients with unilateral absence of wolffian duct structures: further evidence of the mesodermal origin of the prostatic central zone. J Urol 160: 2126–2129.

Berkovitch, M, Hunt D, Koelln N, Morgan T, Collins AL, Faihurst J, Robertson SP, Douglas AGL, Gorlery A. 2017. Quantification of transmission risk in a male patient with a FLNB mosaic mutation causing Larsen syndrome: implications for genetic counseling in postzygotic mosaicism cases. Hum Mutat 38: 1360–1364.

Campbell JM, Stewart JR, James RA, Lupski JR, Stankiewicz P, Olofsson P, Shaw CA. 2014. Parent of origin, mosaicism, and recurrence risk: probabilistic modeling explains the broken symmetry of transmission genetics. Am J Hum Genet 95: 345–359.

Chen Z, Moran K, Richards-Yutz J, Toonen E, Gerhardt D, Ganguly T, Shields GJ, Ganguly A. 2014. Enhanced sensitivity for detection of low-level germline mosaic RB1 mutations in sporadic retinoblastoma cases using deep semiconductor sequencing. Hum Mutat 35: 384–391.

Dal GM, Erguner B, Sagiroglu MS, Yuksel B, Onat OE, Alkan C, Orzetlik T. 2014. Early postzygotic mutations contribute to de novo variation in a healthy monozygotic twin pair. J Med Genet 51: 455–459.

De S, 2011. Somatic mosaicism in healthy human tissues. Trends Genet 27: 217–223.

Depienne C, Arzimanoglou A, Trouillard O, Fedirko E, Baulac S, Saint-Martin C, Roberge M, Dravet C, Nabbout R, Baulac M, et al. 2006. Parental mosaicism can cause recurrent transmission of SCN1A mutations associated with severe myoclonic epilepsy of infancy. Hum Mutat 27: 389.

Depienne C, Trouillard O, Gourfinkel-An I, Saint-Martin C, Bouteiller D, Graber D, Barthez-Carpentier M-A, Gautier A, Villeneuve N, Dravet C, et al. 2010. Mechanisms for variable expressivity of inherited SCN1A mutations causing Dravet syndrome. J Med Genet 47: 404–410.

Dou Y, Yang X, Li Z, Wang S, Zhang Z, Ye AY, Yan L, Yang C, Wu Q, Li J, et al. 2017. Post-zygotic single-nucleotide mosaics contribute to the etiology of autism spectrum disorder and autistic traits and the origin of mutations. Hum Mutat 38: 1002–1013.

Fischbach GD, Lord C. 2010. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neurosci 68: 192–195.

Fisher RA. 1922. On the dominance ratio. Proc R Soc Edinb 42: 321–341.

Gao JJ, Fan XR, Hu J, Ma L, Wu M, Shao YL, Barton SA, Woodruff RC, Zhang YP, Fu YX. 2011. Highly variable recessive lethal or nearly lethal mutation rates during germ-line development of male Drosophila melanogaster. Proc Natl Acad Sci USA 108: 15914–15919.

Gilbert SF. 2003. Paraxial and intermediate mesoderm. In Developmental Biology, pp. 465–490. Sinauer Associates, Inc., Sunderland, MA.

Goriely A, McGrath JJ, Hultman CM, Willkow A, Malaspina D. 2013. “Selfish spermatogonial selection”: a novel mechanism for the association between advanced paternal age and neurodevelopmental disorders. Am J Psychiatry 170: 599–608.

Hardy K, Handside AH, Winston RM. 1989. The human blastocyst: cell number, death and allocation during late preimplantation development in vitro. Development 107: 597–604.

Hardy K, Spanos S, Becker D, Iannelli P, Winston RM. 1993. From cell death to embryo arrest: mathematical models of human preimplantation embryo development. Proc Natl Acad Sci USA 90: 1655–1660.

Huang AY, Xu X, Ye AY, Wu Q, Yan L, Zhao B, Yang X, He Y, Wang S, Zhang Z, et al. 2014. Postzygotic single-nucleotide mosaics in whole-genome sequences of clinically unremarkable individuals. Cell Res 24: 1311–1327.

Huang AY, Zhang Z, Ye AY, Dou Y, Yan L, Yang X, Zhang Y, Wei L. 2017. MosaicHunter: accurate detection of postzygotic single-nucleotide mosaicism through next-generation sequencing of unpaired, trio, and paired samples. Nucleic Acids Res 45: e76.

Huang AY, Yang X, Wang S, Zheng X, Wu Q, Ye AY, Wei L. 2018. Distinctive patterns of postzygotic single-nucleotide mosaics in healthy individuals revealed by genome-wide profiling of multiple organs. PLoS Genet 14: e1007395.

Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumn M, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, et al. 2014. The contribution of de novo coding mutations to autism spectrum disorder. Nature 515: 216–221.

Jones AC, Sampson JR, Chendle JP. 2001. Low level mosaicism detectable by DHPLC but not by direct sequencing. Hum Mutat 17: 233–234.

Ju YS, Martincocorea I, Gerstung M, Petljak M, Alexandrov LB, Rahbari R, Wedge DC, Davies HR, Ramakrishna M, Fillion A, et al. 2017. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. Nature 543: 714–718.

Kimura M. 1954. Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. Genetics 39: 280–295.

Kimura M. 1955. Solution of a process of random genetic drift with a continuous model. Proc Natl Acad Sci USA 41: 144–150.

Kono M, Suga Y, Akashi T, Ito Y, Takeichi T, Muro Y, Akiyama M. 2017. A child with epidermolytic ichthyosis from a parent with epidermolytic

950 Genome Research

www.genome.org

Ye et al.
nevus: risk evaluation of transmission from mosaic to germline. J Invest Dermatol 137: 2024–2026.

Lupski JR. 2013. Genome mosaicism—one human, multiple genomes. Science 341: 588–589.

Meng H, Xu HQ, Yu L, Lin GW, He N, Su T, Shi YW, Li B, Wang J, Liu XR, et al. 2015. The SCN1A mutation database: updating information and analysis of the relationships among genotype, functional alteration, and phenotype. Hum Mutat 36: 573–580.

Moran PA. 1958. Random processes in genetics. Math Proc Camb Philos Soc 54: 60.

Mottla GL, Adelman MR, Hall JL, Gindoff PR, Stillman RJ, Johnson KE. 1995. Fertilization and early embryology: Lineage tracing demonstrates that blastomeres of early cleavage-stage human pre-embryos contribute to both trophectoderm and inner cell mass. Hum Reprod 10: 384–391.

Pansky B. 1982. 25. Germ layers and their derivatives. In Review of medical embryology, p. 25. Embryome Sciences, Inc., Alameda, CA.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. Genetics 192: 1065–1093.

Poduri A, Evrony GD, Cai X, Walsh CA. 2013. Somatic mutation, genomic variation, and neurological disease. Science 341: 1237758.

Priest JR, Gawad C, Kahlig KM, Yu JK, O’Hara T, Boyle PM, Rajamani S, Clark MJ, Garcia STK, Ceresnak S, et al. 2016. Early somatic mosaicism is a rare cause of long-QT syndrome. Proc Natl Acad Sci 113: 11555–11560.

Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, Dominiczak A, Morris A, Porteous D, Smith B, et al. 2016. Timing, rates and spectra of human germline mutation. Nat Genet 48: 126–133.

Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, et al. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468: 1053–1060.

Strachan T, Read AP. 1999. Human molecular genetics, 2nd ed. Wiley-Liss, New York.

Takahashi S, Matsutuji M, Yonee C, Tsuru H, Sano N, Oguni H. 2017. Somatic mosaicism for a SLC2A1 mutation: implications for genetic counseling for GLUT1 deficiency syndrome. Clin Genet 91: 932–933.

Terracciano A, Trivisano M, Cusmai R, De Palma L, Fusco L, Compagnucci C, Bertini E, Vigevano F, Specchio N. 2016. PCDH19-related epilepsy in two mosaic male patients. Epilepsia 57: e51–e55.

Uven O, Turner J, Mathieson I, de Groot N, Bonnrop R, McVean G. 2014. Strong male bias drives germline mutation in chimpanzees. Science 344: 1272–1275.

Wright S. 1931. Evolution in Mendelian populations. Genetics 16: 97–159.

Wright S. 1939. Statistical genetics in relation to evolution. Expo Biométrie Stat Biol 87: 430–431.

Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, McMichael JF, Schmidt HK, Yellapantula V, Miller CA, et al. 2014. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med 20: 1472–1478.

Xu X, Yang X, Wu Q, Liu A, Yang X, Ye AY, Huang AY, Li J, Wang M, Yu Z, et al. 2015. Amplicon resequencing identified parental mosaicism for approximately 10% of “de novo” SCN1A mutations in children with Dravet syndrome. Hum Mutat 36: 861–872.

Yang X, Liu A, Xu X, Yang X, Zeng Q, Ye AY, Yu Z, Wang S, Huang AY, Wu X, et al. 2017. Genomic mosaicism in paternal sperm and multiple parental tissues in a Dravet syndrome cohort. Sci Rep 7: 15677.

Yuen RK, Alipanahi B, Thiruvahindrapuram B, Tong X, Sun Y, Cao D, Zhang T, Wu X, Jin X, Zhou Z, et al. 2016. Genome-wide characteristics of de novo mutations in autism. NPJ Genom Med 1: 16027.