Parent of Origin, Mosaicism, and Recurrence Risk: Probabilistic Modeling Explains the Broken Symmetry of Transmission Genetics

Ian M. Campbell,[1] Jonathan R. Stewart,[2] Regis A. James,[3] James R. Lupski,[1,4,5,6] Paweł Stankiewicz,[1,7] Peter Olofsson,[8] and Chad A. Shaw[1,2,9,*]

Most new mutations are observed to arise in fathers, and increasing paternal age positively correlates with the risk of new variants. Interestingly, new mutations in X-linked recessive disease show elevated familial recurrence rates. In male offspring, these mutations must be inherited from mothers. We previously developed a simulation model to consider parental mosaicism as a source of transmitted mutations. In this paper, we extend and formalize the model to provide analytical results and flexible formulas. The results implicate parent of origin and parental mosaicism as central variables in recurrence risk. Consistent with empirical data, our model predicts that more transmitted mutations arise in fathers and that this tendency increases as fathers age. Notably, the lack of expansion later in the male germline determines relatively lower variance in the proportion of mutants, which decreases with paternal age. Subsequently, observation of a transmitted mutation has less impact on the expected risk for future offspring. Conversely, for the female germline, which arrests after clonal expansion in early development, variance in the mutant proportion is higher, and observation of a transmitted mutation dramatically increases the expected risk of recurrence in another pregnancy. Parental somatic mosaicism considerably elevates risk for both parents. These findings have important implications for genetic counseling and for understanding patterns of recurrence in transmission genetics. We provide a convenient online tool and source code implementing our analytical results. These tools permit varying the underlying parameters that influence recurrence risk and could be useful for analyzing risk in diverse family structures.

Introduction

New mutations are the sole source of disease risk for genetic disorders that eliminate reproductive fitness and for lethal alleles that can only exist in a mosaic state. Likewise, new mutations account for approximately one-third of disease risk in severe X-linked recessive conditions that diminish reproduction. In some instances, these new mutations are mitotic in origin (they arise during embryologic development of a parent) and are present in a low-level mosaic state. Such mutations can include single-nucleotide variations (SNVs), indels, nonrecurrent copy-number variations (CNVs), and other nonrecurrent copy-number-neutral structural variations.[1] Importantly, these mutations can be present in the germline of parents and can be potentially recurrently transmitted to future offspring.[2–4] Unexpected recurrences can occur, as evidenced by multiple affected children harboring the same apparently de novo variation. The birth of a single child with a severe genetic disease presents considerable psychological, social, and economic challenges; consequently, recurrence of the same disorder in a second child is a situation many couples prefer to avoid.[5] Families who have had children affected by apparently de novo mutations can therefore benefit from well-informed risk counseling to make reproductive choices and plan prenatal care for additional pregnancies. These recurrence-risk estimates are an important aspect of the health care provided to such couples, particularly in severe and highly penetrant genetic disorders for which medical therapy remains limited.

Geneticists commonly use the value of $<1\%$ to estimate the risk of recurrence for simplex de novo mutations[6] to be transmitted to additional pregnancies. However, consideration of the literature shows that this risk assessment is often inconsistent with empirical risk for some specific disorders,[7] particularly those caused by mutations in genes located on the X chromosome.[8,9] These examples provide insight into understanding exceptions to rarity of recurrence of apparently de novo mutations: males affected by X-linked recessive conditions necessarily harbor mutations on the chromosome inherited from their mothers. This maternal bias stands in contrast to the observations that most new mutations arise in the paternal lineage and that the risk of de novo mutation increases with paternal age.[10,11] This paternal bias is broadly consistent with the mitotic origin of many de novo mutations and the additional mitoses experienced by germ cells as fathers age.

We hypothesized that sexual dimorphism in gametogenesis might underlie the juxtaposition of these contrasting biases in higher recurrence risk for X-linked disease and the increased paternal origin of most de novo transmitted

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [2]Department of Statistics, Rice University, Houston, TX 77005, USA; [3]Department of Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030, USA; [4]Texas Children's Hospital, Houston, TX 77030, USA; [5]Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; [6]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; [7]Institute of Mother and Child, Warsaw 01-211, Poland; [8]Mathematics Department, Trinity University, San Antonio, TX 78212, USA; [9]Department of Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030, USA
*Correspondence: cashaw@bcm.edu
http://dx.doi.org/10.1016/j.ajhg.2014.08.010.

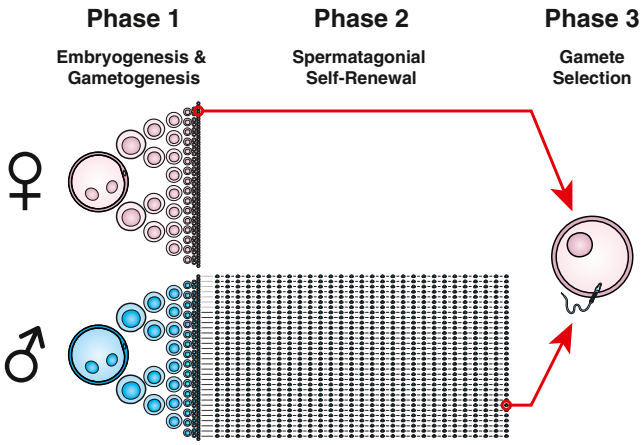

Figure 1. Stochastic-Process Model of Sexual Dimorphisms during Gametogenesis
Phase 1: both males and females experience a stochastic exponential cell-expansion phase modeling embryogenesis and germ cell proliferation. Mutations can arise in any cell division, and if they persist in the clonal lineage, they could ultimately be available to be transmitted to the next generation.
Phase 2: in males, expansion is followed by a stochastic but nonexpanding self-renewal process modeling spermatogenesis.
Phase 3: a single sperm and egg are randomly sampled after meiosis to fertilize an offspring. Adapted from Campbell et al.[4] with permission.

mutations. To address this hypothesis, we developed a comprehensive, flexible mathematical model that describes the emergence of new transmitted variants. We show how recurrence risk can be computed on the basis of conditional probability analyses applied to these models in the context of observed affected offspring. These analyses give a comprehensive picture of the emergence of de novo variation and a systematic framework for quantitative analysis of recurrence. The main conclusion of our work is that the parent of origin and the presence of parental somatic mosaicism are major determinants of recurrence risk.

Material and Methods

Models of Mutation and Germline Development

For our investigations, we utilized multitype Galton-Watson processes to model gametogenesis. These approaches are well established in probability theory and have been used for over a century; however, they have seldom been used in statistical genetics. Previous studies of mosaicism led us to hypothesize that this modeling framework could be useful in the analysis of sexual dimorphisms in gametogenesis.[4] Our model of gametogenesis is composed of three stages (Figure 1). We can optionally include a fourth initial stage that allows stochastic exponential growth without mutation to exclude extremely early embryologic mutations that potentially cause the somatically mosaic parent to be affected. In the first stage of our main model, we consider clonal expansion that results in the initial germ pool during embryogenesis. This expansion phase is parameterized by three variables: the doubling rate of wild-type cells (p), the doubling rate of mutant cells (q), and the per-mitosis mutation rate (λ_1). In the second male-specific stage, we consider the self-renewing process of spermatogenesis. This phase is relatively stable in terms of the total population size and is parameterized by five variables: the doubling rate of wild-type spermatogonial cells (α), the self-renewal rate of wild-type cells (β), the doubling rate of mutant spermatogonial cells (γ), the self-renewal rate of mutant cells (ξ), and the per-mitosis mutation rate (λ_2). The final phase is meiosis, where diploid mutant cells give rise to haploid mutant gametes at a rate of 50% of the pool of mutant cells. We then select a single gamete from each parent to determine transmission.

We previously developed exact formulas for analyzing single-phase Galton-Watson models.[12] Here, we extended these sampling formulas to encompass the multistage process (see Appendix A). In brief, this approach requires composition of the output of each prior phase and the subsequent phase. We determined an exact integral expression for the mean and variance of the proportion of cells with mutations at a particular locus within a parent's germ pool on the basis of the probability generating functions of the composite process. We also developed a recursive computational scheme to numerically determine the required integrands. Mathematica code implementing the formulas determined in Appendix A is provided on our website.

However, for numbers of mitoses consistent with human development, it is impractical to compute the number of terms in the generating-function expansions. Furthermore, numerical integration approaches based on a finite grid break down well in advance of useful numbers of mitoses. Therefore, we sought an alternative method to determine model properties and recurrence-risk estimates. We developed an exact matrix formulation for the first two moments of the multistage Galton-Watson model (Appendix B). We used these results and Taylor approximations for the moments of functions of random variables (in this case, a proportion) to determine approximations of the mean and variance of the proportion of mutants as a function of paternal age and other model parameters (Appendix B).

Updating the expected proportion of mutants given the observation of a transmitted mutation is essential to determining recurrence risk. We used the definition of conditional probability to determine an expression for the conditional expectation of the proportion of mutants on the basis of the unconditional moments of the proportion. Our results reveal that the conditional expectation of the proportion of mutants given an observed transmission is equivalent to the size-biased mean of the proportion, a more general mathematical result (Appendix C).

We subsequently explored the Beta-Binomial conjugate family[13] as a useful and convenient Bayesian method to determine risk in diverse family structures. We parameterized a Beta distribution for the unconditional mean and variance of the proportion of mutants by using the results from the Galton-Watson model. We observed that the Beta-Binomial model gives exactly the same result as that obtained by the size-biased mean of the proportion of mutants for a single transmission. Therefore, we utilized the Beta-Binomial model as a flexible and effective method to estimate recurrence risk for arbitrary family sizes and the number of affected and unaffected offspring. Notably, unaffected offspring provide different information about the proportion of mutants in the germ pool of the transmitting parent depending on whether or not they inherit the risk haplotype—the chromosome on which the transmitted new mutation occurred. If haplotype information is unavailable, the information for updating the proportion is correspondingly diminished. Inheritance of the nonrisk haplotype conveys almost no information. To incorporate