ESHG

BRIEF COMMUNICATION

On the loss of human sex chromosomes in lymphocytes with age: a quantitative treatment

Reiner A. Veitia [1,2]

Abstract

Increasing levels of aneuploidy of the sex chromosomes in human lymphocytes with age have been noted for several decades. The percentage of chromosome Y loss can reach up to 1.5% or even more, whereas the levels of X0 cells in females can increase up to 5% with age. Here, I propose simple mathematical models of the dynamics of 'normal' sex chromosome loss with age. These exponential models provide more mechanistic insights than linear regressions. They account for the lower incidence of sex chromosome loss in young individuals and its increase with age. Moreover, the exponential models show that aneuploidy of the sex chromosomes provides a selective advantage. As there are no longitudinal data available (for the same individual at different time points), the parameters reported here are average values derived from a population. Hopefully, this study will stimulate further work based on next-generation technologies to obtain better estimates of sex chromosome aneuploidy and of the parameters of the models discussed here.

Introduction

An increasing aneuploidy in human lymphocytes with age has been noted since the 1960s [1], which has been confirmed particularly for the sex chromosomes [2, 3]. The most complete dataset available on the loss human X and Y chromosomes used in situ hybridization to determine the frequency of aneuploidy in both males and females [3]. A total of 1000 lymphocyte interphase nuclei were screened per individual, for 90 females and 138 males of different ages. The percentage of chromosome Y loss was found to rise from about 0 up to 1.5% with age. A plot of the proportion of sex chromosome loss (F) versus age (t) was described by a straight line of the form $F = at + b$. In females, the initial levels of X0 cells were higher and increased up to 5% with age. On the contrary, the frequency of autosomal loss did not change with age. Aneuploidy of the sex chromosomes probably reflects a loss at the level of lymphopoietic stem cells, which is subsequently detected in circulating cells [3].

The loss of the Y chromosome has also been observed in hematological malignancies. A study involving leukaemic or preleukaemic male patients showed that 3.4% had a 45,X0 clone in their bone marrow, a phenomenon called clonal hematopoiesis (CP) [4]. CP is also associated with increased mortality, hematological malignancy, and Y-chromosome loss [5]. Along similar lines, the analysis of elderly men found an association between the loss of the Y chromosome in leukocytes and the risk of mortality both, in general, and from non-hematological cancer [6].

Below, I present mathematical models of the dynamics of 'normal' sex chromosome loss with age. Such mechanistic models are very simple and can thus be used both in research and teaching settings. I hope this analysis will stimulate the production of data based on next generation technologies that will help to obtain a better quantitative estimation of X and Y chromosome aneuploidy.

Methods

As there is no longitudinal data of sex chromosome loss for the same individual, I use published population data (for many individuals with different ages) to test the exponential models developed below. A total of 123 pairs of values (age and proportion of aneuploid cells), representing as many

✉ Reiner A. Veitia
   reiner.veitia@ijm.fr

1   Institut Jacques Monod, Université Paris Diderot, Paris, France

2   Université Paris-Diderot, Paris, France
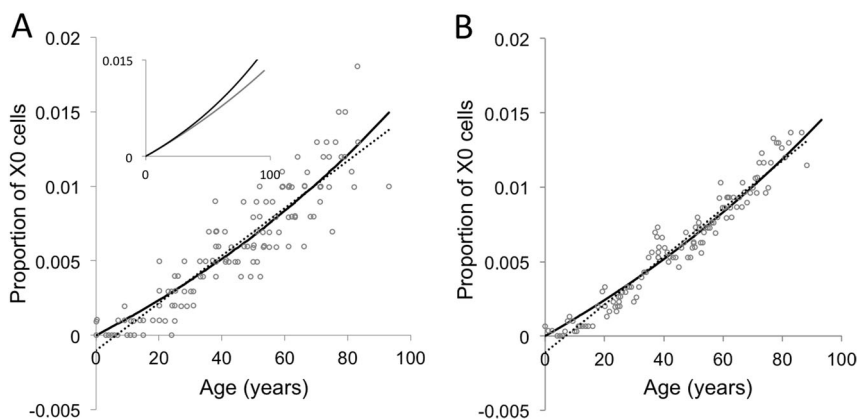
A

B

Fig. 1 a Plot of *F* of Y loss versus age for 123 data points obtained from ref. [3]. The solid line corresponds to the one predicted by the exponential model. The dotted line corresponds to the regression line $F = 0.00015t + 0.001$. The insert corresponds to curves for $k_l = 0.0001$ and $k' = 0.005$ in gray and 0.009 in black. **b** Plot of *F* of Y loss

versus age where each point corresponds to the average of three original data points calculated with a sliding window. This decreases the scattering and is displayed only to help the reader to better appreciate the non-linearity

male individuals, were obtained from ref. [3] for the loss of Y chromosome. Regarding X loss, 89 pairs of values (age and proportion of X0 cells) were obtained from ref. [3]. As data were displayed in graphical form, the numerical values were extracted using the software WebPlotDigitizer at https://automeris.io/WebPlotDigitizer/. Nonlinear curve fitting was performed using the least square methods at http://statpages.info/nonlin.html.

Results and discussion

The dynamics of euploid (*XY*) and Y-aneuploid (*X0*) cell sub-populations can be described by the following equations:

$$\frac{\mathrm{d}XY}{\mathrm{d}t} = k_{XY}XY - \delta_{XY}XY - k_lXY, \tag{1}$$

$$\frac{\mathrm{d}X0}{\mathrm{d}t} = k_lXY + k_{X0}X0 - \delta_{X0}X0, \tag{2}$$

where *XY* and *X0* are the numbers of cells carrying the Y chromosome or not, respectively. The constants k_{XY} and k_{X0} represent (probably stem) cell proliferation rates for the corresponding cell types, k_l stands for the specific rate of loss of the Y (i.e., initial production of X0 cells) and the δ constants encapsulate information on the specific death rates of the corresponding cell sub-populations. As the proportion of *X0* cells is low in apparently normal individuals, we can simplify the mathematical treatment by considering that the population of *XY* cells remains unchanged. This quasi-steady-state assumption is represented by $\mathrm{d}XY/\mathrm{d}t = 0$ in Eq. 1 and is essential to remove parameters which would have been difficult to estimate. A reader not familiar with

mathematical models can directly go to Eqs 3 and 4 and their discussion.

The solution of Eq. 1 implies that $XY = k$ (constant), if we tolerate 2% of error for the number of XY cells at old ages. By virtue of this simplification, Eq. 1' becomes:

$$\frac{\mathrm{d}X0}{\mathrm{d}t} = k_lk + (k_{X0} - \delta_{X0})X0 \tag{1}$$

After simplifying the notation by making $(k_{X0} - \delta_{X0}) = k'$, Eq. 1' can be easily integrated as follows:

$$X0 = ce^{k't} - \frac{k_lk}{k'} \tag{2}$$

As previously shown [3], $X0 = 0$ at $t = 0$. To satisfy this condition, $c = k_lk/k'$. As $XY \gg X0$, we can consider that $XY + X0$ is simply $XY = k$. Thus, the fraction *F* of *X0* cells can be calculated as follows:

$$F = \frac{X0}{XY + X0} = \frac{\frac{k_lk}{k'}\left(e^{k't} - 1\right)}{k} = \frac{k_l}{k'}\left(e^{k't} - 1\right); \tag{3}$$

The model captured by Eq. 3 will be henceforth called the 'exponential model'. It clearly shows that the increase of the proportion of *X0* lymphocytes with age directly depends on their specific rate of appearance k_l and in a more complex way on k', which appears in the denominator of the constant term k_l/k' but also in the exponential term. The positive impact of k' in the exponential term outweighs the negative impact in the denominator. Thus, individuals with higher k' (for the same k_l) will have a greater fraction of aneuploid cells at each time point (Fig. 1).

As mentioned above, *F* versus age can be represented by a 'linear model' $F = at + b$, where *a* is a constant without any clear mechanistic meaning. Moreover, it leads to a

Fig. 2 a Plot of F of X loss versus age for 89 data points obtained from ref. [3]. The solid line corresponds to the one predicted by the exponential model. The dotted line is a regression line $F = 0.0004t + 0.00169$. **b** Plot of F of X loss versus age where each point corresponds to the average of four original data points calculated with a sliding window
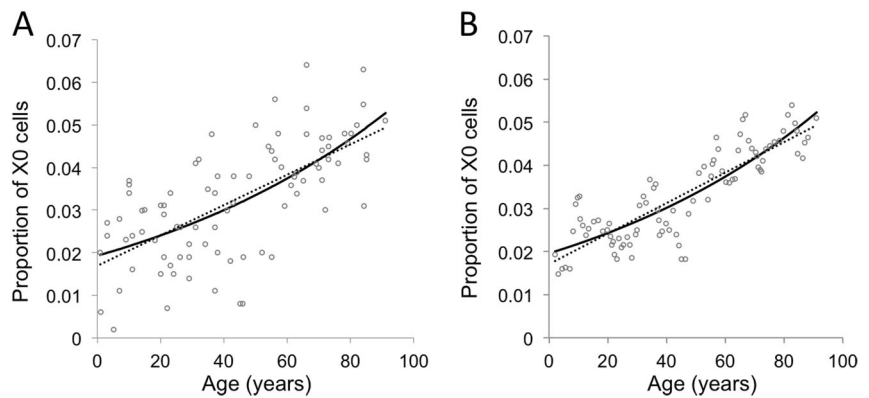
negative value of $b = F_0$ (i.e., at-birth F, Fig. 1). Eq. 3 fits data better than the straight line for the same number of parameters as it provides a smaller sum of the squared errors (i.e., the difference between the observed and predicted values), which is a measure of the goodness of fit. However, its main advantage is to provide mechanistic insights. For instance, if we calculate the slope of the tangent of the curve for $t = 0$ in Fig. 1, we find that it corresponds to $k_l k'/k'$, that is, k_l. This value ($k_l = 0.00011$) is close to $a = 0.00014$ from the linear model, yet nothing in the linear model indicates that $a \approx k_l$. The disparity between the values is explained by the fact that according to the exponential model, the rate of Y loss is slower at young ages and is predicted to increase with time. Thus, a slope of the linear model computed over the whole data set would be higher that than of that of the tangent at $t = 0$. To avoid dealing with the negative value of F_0, which has no physical meaning, we can consider another linear model, $F = at$. In this case, the goodness of fit is even lower than for $F = at + b$. Indeed, the exponential model is 43 times more likely to be correct than $F = at$ (according to the Akaiké Information Content criterion), which is corroborated by a Fisher's-test ($p < 0.002$). The exponential model explains both the observed low incidence of Y chromosome loss in boys and the faster increase with age, as noted in ref. [3].

According to the exponential model, the specific rate of production of X0 stem cells is $k_l = 0.00011$. Whereas the difference between the specific rates of proliferation and death of X0 cells is $k' = 0.0077$. These subtleties (i.e., production of X0 cells and their net proliferation) are absent from the linear model. According to our assumption of steady-state for the XY cells, $k_{XY} - \delta_{XY} \approx k_l = 0.00011$. If we compare this value with $k' = k_{X0} - \delta_{X0} = 0.0077$ for the Y-less cells, we immediately see that losing the Y chromosome provides a selective advantage. The net proliferation capacity X0 cells can be estimated at about 70 (i.e., 0.0077/0.00011) times that of euploid cells. However, the former do not invade the population because they are produced at a very slow pace (k_l) even if, in pathological conditions, their

presence and consequences can be notorious [4–6]. A selective advantage of X0 cells has been suggested to explain why a 46,XX/45,X0 girl diagnosed at 10 years of age showed no evidence of 46,XX cells in her tissues analyzed at 25 [7]. Lifestyle is also likely to have an impact of sex chromosome loss. For instance, smoking is associated with the loss of the Y in blood cells to levels much higher than those discussed above [8].

We can now turn to explore what happens in the case of the loss of one X chromosome. The first thing worth noting is that data in females show that a proportion of X0 cells is already present at very young ages (Fig. 2). Thus, we have to simplify Eq. 2 otherwise, for instance, by neglecting k_l/k' and by tolerating 5% of error regarding the number of XX cells at old ages. Thus, F can be calculated as follows:

$$F = \frac{X0}{XX + X0} = \frac{ce^{k't} - \frac{k_l k}{k'}}{k} = \frac{ce^{k't}}{k} - \frac{k_l}{k'} = F_0 e^{k't} \quad (4)$$

The plot of F versus age has also been described by $F = at + b$ (Fig. 2). Again, Eq. 4 provides a better fit and mechanistic explanations. The tangent of the curve at $t = 0$ is $F = F_0 * k't + F_0$. This translates into $F = 0.0002t + 0.0192$, which is not far from the equation of the regression line $F = 0.0004t + 0.0169$. The explanation for the difference in the slopes is the same as above for the X0 cells in an XY background. It is tempting to estimate the degree of proliferative advantage of X0 cells as done above in an XY context. Although, k_l has disappeared from the most simplified form of Eq. 4, we can still obtain an estimate using a less simplified one. However, such an estimate would be less reliable than that for the loss of the Y. This is due to the number of parameters to be estimated and to the greater dispersion of the data points because of the difficulty in scoring of XX and X0 nuclei due to overlapping FISH signals.

The selective advantage of X0 cells might result from the haploinsufficiency of genes encoding negative regulators of cell proliferation and containing expressed homologs on both sex chromosomes (XY) or its copies (XX) or to the loss of potential Y-specific tumor suppressor genes. In line with

this, it is known that *ZFY* and *UTY* have properties of tumor suppressors and have homologues on the X that escape inactivation [9]. *TMSB4Y* is another candidate tumor suppressor on the Y, which is deleted in male breast cancer [10].

Before closing, it is worth mentioning some of the limitations of this analysis. As there is no longitudinal data available, the parameters obtained are average values derived from a population of different individuals. Current data show great dispersion and interindividual variation is expected. Moreover, the proportion of X chromosome loss being higher than that for the Y leads to less accurate predictions due to our simplifying approximations. Thus, further studies using up-to-date technologies are required to obtain a better appraisal of age-related sex chromosome aneuploidy and its effects.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Jacobs PA, Brunton M, Court Brown WM, Doll R, Goldstein H. Change of human chromosome count distribution with age: evidence for a sex differences. Nature 1963;197:1080–1.
2. Fitzgerald PH, McEwan CM. Total aneuploidy and age-related sex chromosome aneuploidy in cultured lymphocytes of normal men and women. Hum Genet. 1977;39:329–37.
3. Guttenbach M, Koschorz B, Bernthaler U, Grimm T, Schmid M. Sex chromosome loss and aging: in situ hybridization studies on human interphase nuclei. Am J Hum Genet. 1995;57: 1143–50.
4. Herens C, Brasseur E, Jamar M, Vierset L, Schoenen I, Koulischer L. Loss of the Y chromosome in bone marrow cells: results on 1907 consecutive cases of leukaemia and preleukaemia. Clin Lab Haematol. 1999;21:17–20.
5. Zink F, Stacey SN, Norddahl GL, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood. 2017;130:742–52.
6. Forsberg LA, Rasi C, Malmqvist N, et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. Nat Genet. 2014;46:624–8.
7. McCorquodale MM, Bowdle FC. Two pregnancies and the loss of the 46,XX cell line in a 45,X/46,XX Turner mosaic patient. Fertil Steril. 1985;43:229–33.
8. Dumanski JP, Rasi C, Lönn M, et al. Mutagenesis. Smoking is associated with mosaic loss of chromosome Y. Science. 2015;347:81–83.
9. Davoli T, Xu AW, Mengwasser KE, et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell. 2013;155:948–62.
10. Wong HY, Wang GM, Croessmann S, et al. TMSB4Y is a candidate tumor suppressor on the Y chromosome and is deleted in male breast cancer. Oncotarget. 2015;6:44927–40.