

Assessment of mechanism of acquired skewed X inactivation by analysis of twins

Mark A. Vickers, Ewan McLeod, Timothy D. Spector, and Ian J. Wilson

Skewed X-chromosome inactivation in peripheral blood granulocytes becomes more frequent with increasing age, affecting up to half of those over 75 years old. To investigate the mechanisms underlying this phenomenon, X-inactivation profiles in 33 monozygotic and 22 dizygotic elderly twin pairs were studied. Differential methylation-sensitive restriction enzyme cutting at a hypervariable locus in the human androgen receptor gene (*HUMARA*) was studied on purified granulocytes using T cells as controls. A large genetic effect on skewed granulocytic X

inactivation was shown ($P < .05$); heritability was estimated to be 0.68. A minor part (SD .0151 relative allele frequency [ie, larger/smaller] units) of the observed variance is due to experimental error. A further contributor to acquired skewing is stochastic asymmetric stem cell division, which was modeled and shown as unlikely to account for a substantial part of variance. Two monozygotic twin pairs had X-inactivation ratios skewed markedly in opposite directions, evidence for a further stochastic mechanism, suggestive of a single overrepresented clone. In con-

clusion, all 3 suggested mechanisms contribute to acquired X inactivation but the dominant mechanism is genetic selection. The observed proportion of putatively clonal hematopoiesis is similar to the lifetime incidence of hematopoietic stem cell malignancy consistent with the concept that clonal hematopoiesis precedes stem cell malignancy. (Blood. 2001; 97:1274-1281)

© 2001 by The American Society of Hematology

Introduction

One of 2 X chromosomes is randomly inactivated early in female embryonic development,¹ so that it becomes transcriptionally inactive and thereby equalizes the dosage of X-linked genes between males and females. After this period, techniques able to distinguish between the 2 X chromosomes can be used to track the fate of the 2 different cellular populations. Initially distinction was achieved using polymorphic enzyme variants,² but, more recently, the fact that placental mammals methylate the inactive X chromosome and certain restriction enzymes selectively cut the unmethylated allele has allowed more widespread application.^{3,4} In particular, the human androgen receptor locus has been found to be useful due to its high proportion (about 90%) of polymorphism.⁵ In this way, many tumors,⁶ including leukemias,^{7,8} myelodysplasia,⁹ and myeloproliferative disorders,¹⁰⁻¹² have been shown to express only one X chromosome, implying derivation from a single cell.

The process of X inactivation is thought to be initiated by a *cis*-acting element at the X-inactivation center, Xist,¹³⁻¹⁶ first expressed at the 4-cell stage. A genome-wide demethylation occurs between the 8-cell stage and the blastocyst stage, so that Xist expression is expressed randomly with respect to the paternal and maternal X chromosomes. The X inactivation that ensues, before the gastrula stage, is then believed to be random. X inactivation, or lyonization, ratios in tissues then follow a binomial distribution. This theory has allowed estimation of the number of hemopoietic precursor cells at the time of lyonization to be calculated in bone marrow,¹⁷⁻²⁰ with other organs tending to inactivate at different

times.²¹ However, as more experience has been gained it was suspected,²² then demonstrated, that it is a common event occurring in up to half of elderly women.^{20,23-25} Furthermore, this acquired skewing seems to be most marked in the granulocytic fraction, the distribution for elderly T cells being similar to that for younger subjects.

Three mechanisms underlying acquired skewed X inactivation have been proposed. First, a "selective advantage" of one X chromosome over the other, the only nonstochastic mechanism.²⁶ This is a well-described phenomenon in the extreme case where one X chromosome is known to carry a disease gene and appears to be selected against.²⁷ That this mechanism can operate in the absence of an obviously deleterious phenotype has been shown in cats, where the X chromosome of the Geoffroy cat has a selective advantage over that of the domestic cat.²⁸ Furthermore, a genetic influence on X-chromosome representation in the blood of elderly monozygotic human twins has been reported.²⁹ Second, because it is believed that stem cell division may stochastically result in zero, 1, or 2 daughter stem cells,^{25,30} certain stem cells may become overrepresented or correspondingly underrepresented by chance. This effect may be accentuated by stem cell depletion.³⁰ Third is the expansion of a clone derived from a single stem cell or stem cell precursor with a "proliferative advantage."

In this paper, the importance of the first proposed mechanism is assessed by reporting the skewing of X inactivation in monozygotic and dizygotic twins. If genetic selection were an important

From the Department of Haematology, Medicine and Therapeutics, University of Aberdeen, Foresterhill, Aberdeen, UK; the Twin Research and Genetic Epidemiology Unit, St Thomas's Hospital, London, UK; and the Department of Mathematical Sciences, Meston Building, University of Aberdeen, Old Aberdeen, UK.

Submitted April 4, 2000; accepted October 6, 2000.

Supported by the Aberdeen Royal Infirmary Leukaemia Research Fund. The Twin Research Unit receives grants from the Arthritis and Rheumatism campaign, the Wellcome Trust and Gemini Genomics Ltd.

Reprints: Mark A. Vickers, Department of Haematology, Medicine and Therapeutics, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD, UK; e-mail: m.a.vickers@abdn.ac.uk.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2001 by The American Society of Hematology

mechanism, then the direction and magnitude of acquired skewing in granulocytes compared to T cells should be more similar in monozygotic twins than expected by chance alone. We assess stochastic mechanisms by both measuring the variance of the difference within twin pairs beyond experimental error and also by modeling stem cell division mathematically. Finally, by deducting the contributions of the first 2 mechanisms, we assess the relative importance of the third.

Materials and methods

Subjects

Following informed consent, 5 to 10 mL blood was obtained from subjects in the St Thomas's twin registry; 120 women comprising 33 monozygotic and 27 dizygotic pairs were recruited. The mean ages (SD; range) were 61 (6.0; 51-73) and 63 (3.4; 55-72) years, respectively. Zygosity of twins was documented by an initial validated questionnaire with 95% accuracy, followed by genetic fingerprinting in ambiguous cases using microsatellite analysis at 6 to 7 polymorphic loci.

Cellular fractionation

After overnight transport, peripheral blood granulocytes and mononuclear cells (PBMC) were isolated by density gradient centrifugation on Histopaque 1077 and 1119 (Sigma, Poole, United Kingdom) according to the manufacturer's instructions. Granulocyte fractions were assessed morphologically and were typically more than 99% pure. T cells were isolated from the mononuclear cell fraction predominantly by E-rosetting, but anti-CD2-coated magnetic beads (Dyna, Oslo, Norway) were also used according to the manufacturer's instructions. E-rosetting was performed by diluting sheep blood 4:1 into Alsever solution, buffy coat depleting, then washing 3 times in physiologically buffered saline or Hanks buffered salt solution before resuspension in RPMI 10 medium at 2%. Heat-inactivated fetal calf serum (2 mL) and 2 mL sheep red blood cells were added to 1 mL PBMC (1.10^7 cells/mL) in RPMI 10 medium. The mixture was incubated for 10 minutes at 37°C, then on ice for 1 to 16 hours. The mixture was then centrifuged through 10 mL Histopaque 1077 at 900g. RPMI 10 medium (45 mL) was added to the pellet and the red cells lysed with 1 mL water. The T cells were then spun down and washed before DNA extraction. T-cell fractionation was assessed morphologically and granulocytic contamination was always less than 5%. Two samples were assessed by flow cytometry using anti-CD45 and anti-CD2, coupled with phycoerythrin and fluorescein isothiocyanate, respectively, and were more than 90% CD2⁺.

X-chromosome inactivation analysis

DNA was extracted from granulocytes and T cells using the Nucleon kit (Scotlab, Coatbridge, United Kingdom).

Analysis of the X-inactivation status at the HUMARA locus was performed according to Allen and coworkers⁵ with slight modifications. DNA (2 μ g) was incubated with 20 U *Hpa*II (MBI Fermentas, Vilnius, Lithuania) in a 30- μ L reaction; then 3 to 6 μ L of the restricted product was subject to polymerase chain reaction (PCR) in 10 mM Tris HCl, pH 8.8, 50 mM KCl, .08% Nonidet-40 using 1.5 U Taq DNA polymerase (MBI Fermentas) on a Techne Progene (Cambridge, United Kingdom) thermal cycler. Primer sequences were 5'-GCT GTG AAG GTT GCT GCT CCT CAT-3' and 5'-TCC AGA ATC TGT TCC AGA GCG TGC-3'. Cycling conditions were 120 seconds at 40°C, 300 seconds at 95°C: 28 \times 45 seconds at 95°C, 30 seconds at 60°C, 30 seconds at 72°C. The former primer was labeled with 5' FAM (Oswel, Southampton, United Kingdom). PCR products were analyzed on 4% denaturing acrylamide gels using an Applied Biosystems Prism 377 (PerkinElmer, Foster City, CA). Band quantitation was performed using the peak values defined by ABI software.

Data were taken as a ratio of peak heights of the longer to shorter alleles. No correction was performed for efficiency of PCR of the 2 alleles using undigested samples because we found that the degree of methylation of

each allele was inversely correlated with efficiency of amplification. This effect will be analyzed in a subsequent paper (in preparation).

Statistical analysis

Ideally, a summary measure of the direction and magnitude of acquired skewing in granulocytes away from T cells would be used to analyze concordance between twin pairs. However, due to the underlying distribution of T-cell values, no such unbiased measure exists. To illustrate this point, consider whether $G = 0.7$ versus $T = 0.5$ exhibits a greater or lesser degree of skewing than $G = 0.2$ versus $T = 0.1$. Instead, we analyze the concordance of allele ratios for each cell type.

The statistical analyses presented here are of 3 forms: nonparametric randomization tests, Pearson or Spearman correlations on both untransformed and transformed data, and Bayesian inference Using Gibb's Sampling using the BUGS program.³¹

For randomization tests, a statistic is calculated from the data and is then compared to a large number of such statistics calculated from randomly matched pairs of individuals. These were generated through permutation, by selecting one member of each twin pair, and matching unselected twins at random. The *P* value of the test is the proportion of permuted values more extreme than our observed value.

We can measure the direction of drift of granulocytes from T cells by subtracting the observed proportion of the upper allele of the granulocytes from the proportion in the T cells. The degree of drift can then be compared in the twin pair by subtracting this measurement of drift in one of the pairs from the other. In this case, if the degree of acquired skewing were the same in each of the twin pairs then the difference would be 0. If they were different, and we ignore the sign by taking the modulus, that is $|x|$, then the value would be greater. In summary, $d = |(t_1 - g_1) - (t_2 - g_2)|$.

To make the data amenable to analysis using normal distributions, the natural logarithm of the ratio of the height of the longer, *hl*, and shorter, *hs*, allele—often called the logit transformation— $r = \text{logit}(p) = \log[p/(1-p)] = \log(hl/hs)$, as $hs = \text{constant}(1-hl)$.

The BUGS package is designed for complex problems with the potential to make inferences about many unknown parameters. The program does not output point estimates; rather it generates the entire joint distribution of the parameters of interest. This is in the form of large (typically 10 000 here) numbers of samples from which we can recreate the distribution or calculate statistics such as the median or mean as point estimates. The structure of these models can be described as Directed Acyclic Graphs. We used 2 models here: one for the joint estimation of experimental errors and the variance covariance matrix, described in this section, and one to estimate stem cell numbers, described in a later section. Further details of the models and code are available from our Web site, <http://www.maths.abdn.ac.uk/~ijw/BUGS>.

For many of the twin pairs we have results from 2 PCR reactions and 3 gel runs, for others 2 gel runs and either one or 2 PCR reactions. This gave us the opportunity to investigate the relative errors in different parts of the genetic analysis: digestion, PCR, gel electrophoresis, and quantification. The errors arising from the quantification of the gel peaks were very small (data not shown). To model these inferences we used a 2-stage process. If the logit of the allele frequency for a cell type in an individual is *r*, then the logit frequency after extraction and the PCR reaction is $r_{\text{pcr}} = r + e_{\text{pcr}}$, where e_{pcr} is a normally distributed error term with mean 0 and SD σ_{pcr} .

The running of the gel and quantification of bands was also assumed to produce an error, e_{run} , which is also normally distributed with mean 0 and SD σ_{run} . Thus the observed logit relative allele frequency $r_{\text{obs}} = r + e_{\text{pcr}} + e_{\text{run}}$.

BUGS allowed us to take account of these errors and estimate their SDs, assuming independence between errors. BUGS treats the true values of *r* and r_{pcr} as unobserved random variables, and samples from the joint distribution with the error SDs conditional on the observed set of r_{obs} .

The logit T-cell relative allele frequencies and logit granulocytic relative allele frequencies, *r*, were then assumed to come from a multivariate normal distribution with mean 0, with an unknown variance-covariance matrix that could be different for both monozygotic twin pairs and dizygotic twin pairs. BUGS sampled from the posterior distribution of

covariance matrices given the data. The correlation could then be calculated from these matrices.

We minimized dependence on the estimates of prior parameters by using very weakly informative (flat) priors. The error terms for PCR and gel runs had independent gamma priors for the precision (1/variance) with mean 1 and variance 1000. The prior precision matrix had a Wishart prior with diagonal elements 1 and off-diagonal elements 0 giving prior correlations between twin pairs with mean 0, and variances with a large range and order of magnitude 1.

Mathematical modeling of stem cell division to explain acquired skewing

Assuming a constant number of stem cells, when a stem cell divides, the 2 daughter cells may both become stem cells. Alternatively, one or neither may become stem cells, instead differentiating or dying. Therefore, the progeny of any single stem cell may become overrepresented or underrepresented in later generations by chance and it has been suggested that this mechanism of asymmetric stochastic stem cell division might explain all or some of acquired X inactivation.^{19,23} We propose that this mechanism is equivalent to random genetic drift, which has been analyzed extensively by population geneticists.³² If a steady-state population of organisms is considered, where 2 alleles are represented with no selective pressure associated with either, the proportion of the population with either allele can similarly fluctuate with time by chance alone. If 2 organisms/stem cells are taken at random, they may both have arisen from a single organism/stem cell that has just divided or alternatively 2 different organisms/stem cells. With N organisms/stem cells and no selection, the probability that they have not come from a single cell is $1/N$. The probability that after g generations they still do not share a common ancestor is

$$\left(1 - \frac{1}{N}\right)^g.$$

This probability of nonrelatedness is equal to $1 - F$, where F is the probability of identity, the probability that genes (here cells) share a common ancestral allele more recently than at some reference in the past. The geometric distribution for nonidentity is well approximated by an exponential distribution, giving $F = 1 - \exp(-g/N)$, where g is the average number of cell divisions and N is the effective population size (which is the actual population size divided by the variance in the number of offspring, which in a purely random model is one). It can be shown that the beta distribution is an exact model for the island model of genetic drift,³³ and is a good approximation when the initial allele frequency is 0.5 and there is no significant fixation, that is, allele proportions of 0 or 1. Our and others' data appear to fit these assumptions. The distribution of allele frequencies follows approximately a beta distribution with parameters $\alpha = (1/F - 1)p$ and $\beta = (1/F - 1)(1 - p)$, where p is the initial allele frequency.³²

We measure the drift of the allele frequency of T cells away from 0.5, and the drift of granulocytes away from T-cell frequencies using g , N , and the beta distribution. The model assumes that the amount of drift is different for each twin pair and estimates the mean and variance of the individual F s. Priors for PCR- and gel-run errors were the same as for the multivariate-normal model, described in the last section. Details of this model are available from our Web site, <http://www.maths.abdn.ac.uk/~ijw/BUGS>.

Results

A total of 120 individuals comprising 33 monozygotic and 27 dizygotic pairs were analyzed. Five dizygotic pairs gave technically unsatisfactory results, mainly because of substandard cellular fractionation in one of the 4 fractions required for each pair, and were excluded from subsequent analyses. Four monozygotic and 3 dizygotic pairs were both monoallelic or had alleles that were so close in size that it was felt that they could not be reliably distinguished. In one of the dizygotic pairs, only one of the pairs

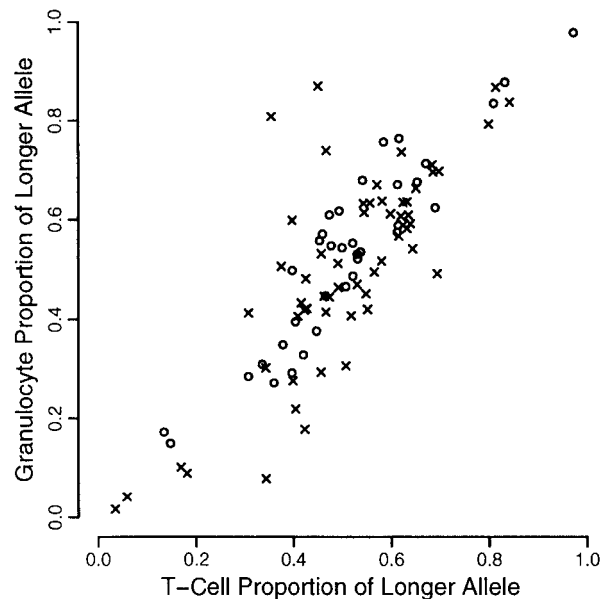


Figure 1. Scatterplot of proportion of longer allele for granulocytes versus T cells. Data from monozygotic twins are marked with crosses and those from dizygotes with open circles.

was monoallelic. These rates are in accordance with previous experience at this locus. All such findings were checked by PCR of undigested samples. These exclusions left 29 monozygotic and 18 dizygotic pairs suitable for statistical analysis. The data are illustrated in Figures 1, 2, and 3. As referred to above, we were unable to develop a completely satisfactory measure of acquired skewing but illustrate a simple measure (allele ratio in granulocytes-allele ratio in T cells) in Figure 4.

Randomization statistics

The randomization and simple correlation tests were performed on the mean of the results from each individual; no account was taken of experimental error. The statistic d was calculated, using the first

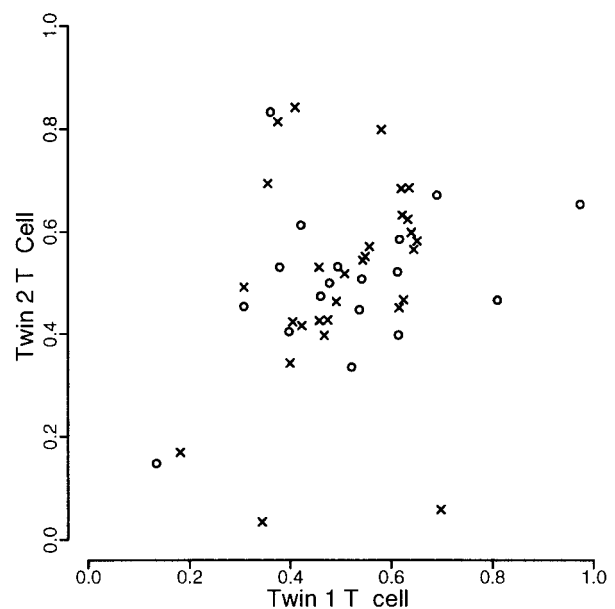


Figure 2. Scatterplot of proportion of longer allele of T-cells within twin pairs. Data from monozygotic twins are marked with crosses and those from dizygotes with open circles.

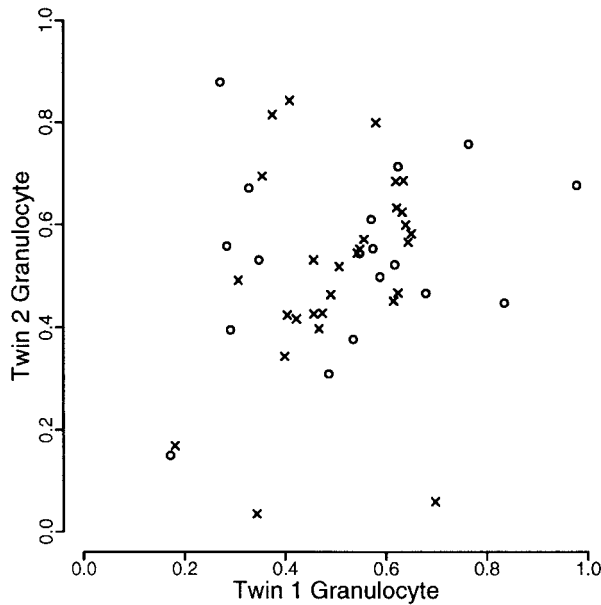


Figure 3. Scatterplot of proportion of longer allele of granulocytes within twin pairs. Data from monozygotic twins are marked with crosses and those from dizygotes with open circles.

equation, to be 2.67 for monozygotic pairs and 1.4 for dizygotic pairs ($P < .003 \pm .0005$ and $< .3 \pm .0045$, respectively).

Simple correlation

The major effect is that the allele ratio of granulocytes is highly correlated with that from T cells within any individual (Pearson correlation coefficient 0.87). The Pearson correlation coefficients (95% confidence intervals) between the untransformed allele ratios within each twin pair are 0.24(−0.10-0.59) and 0.52(0.26-0.79) for the T-cell and granulocytic fractions, respectively, with those for dizygotic twins 0.38(−0.02-0.77) and 0.25(−0.19-0.68), respec-

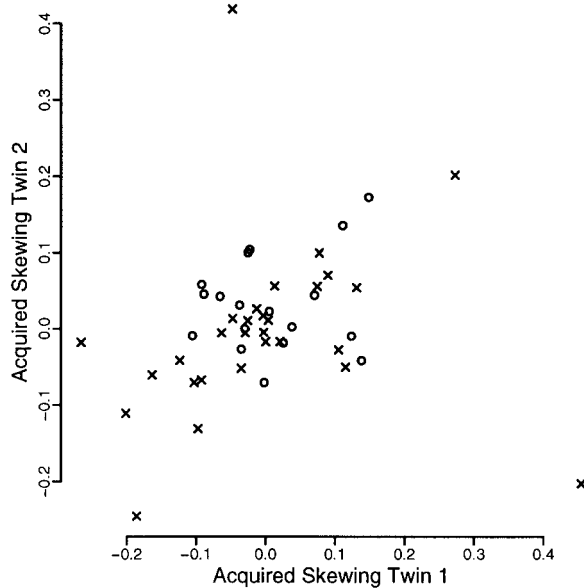


Figure 4. Scatterplot of measure of acquired skewing within twin pairs. The measure of acquired skewing is the proportion of the longer allele seen after restriction in the granulocytic fraction minus that in the T-cell fraction. Data from monozygotic twins are marked with crosses and those from dizygotes with open circles.

tively. Corresponding Spearman (nonparametric) coefficients are 0.28, 0.46, 0.24, and 0.20, respectively.

For subsequent analyses the mean of the logit (h/h_s) was considered; modeling experimental error as assessed by duplicate measurements is considered in a later section. The Pearson correlations (95% confidence intervals) between the T-cell and granulocytic values of monozygotic twins are 0.21(−0.14-0.56) and 0.55(0.3-0.81), respectively, with those for dizygotic twins 0.39(0.00-0.78) and 0.25(−0.19-0.68), respectively.

The only statistically significant result is that for granulocytes within the monozygotic twin pairs. These results give an indication of the association between the twin pairs but take no account of any experimental error.

Measuring experimental error

In units of relative allele frequency (ie, longer/shorter), the estimated SDs are given in Table 1, based on runs of length 10 000 from BUGS. These indicate that levels of experimental error are low compared to other sources of variation in the models. The errors of the 2 stages of measurement appear to have approximately the same magnitude and have considerable overlap. This hierarchical model for relative allele frequencies is used in subsequent analyses unless otherwise stated.

Association between allele frequencies in different twins

The means of all the correlations, illustrated in Figure 5, are more than 0, indicating a positive correlation. Only the granulocytic allele frequencies for monozygotic twins have a negligible proportion of the distribution less than 0. For the monozygotic T cells, dizygotic T cells, and granulocytes, about 10% of the distribution is below 0, which suggests some congenital effect also.

The genetic component can be expressed as the ratio of additive genetic variance to total phenotypic variance (h^2). In the case of twins, this can be expressed as $h^2 = 2(r_{MZ} - r_{DZ})$, where r_{MZ} and r_{DZ} are the intraclass correlations for monozygotic and dizygotic twins, respectively.³⁴ This approach estimates the contribution to variance of the genetic component as $2 \times (0.53-0.19) = 0.68$ for neutrophils and zero for T cells (Table 2). Similar estimates were obtained using variance-components methodology³⁴ (details not shown). In addition, the fact that $r_{MZ} > 2 \cdot r_{DZ}$ suggests the possibility of a dominant effect,³⁵ although the numbers are too small to confirm this.

Mathematical modeling of stem cell division to explain acquired skewing

We use prior distributions for N and G , which are shown in Figure 6 as dotted lines: a distribution for N , which is centered around 10^6 but with values from 10^5 to 10^7 ,³⁶⁻³⁸ and a distribution for g , which had a mean of 50 but allowed values from 0 to 400.³⁹ The inferred distributions are well away from the prior values, indicating that the random drift model is insufficient to explain the amount of drift

Table 1. Analysis of measurement error

	Estimated SDs			
	Percentiles			
	Mean	2.5%	Median	97.5%
Extraction and PCR	0.0083	0.0056	0.0081	0.012
Gel run	0.0068	0.0014	0.0067	0.013

Estimated SDs are given in units of relative allele frequency.

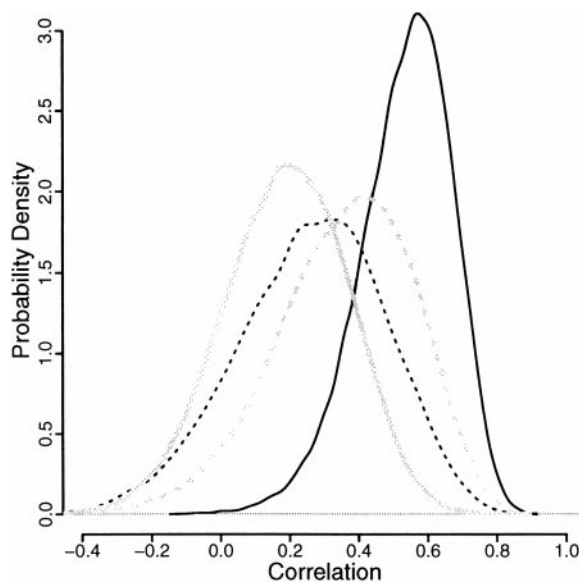


Figure 5. Relative likelihood curves of the correlations for granulocytic and T-cell allele frequencies after logit transformation. Results from granulocytes are in solid lines, those from T cells are in broken lines; both are split by monozygotic (dark lines) and dizygotic (light lines) twins.

seen in these allele frequencies. In other words, only an unrealistically low number of stem cells and/or an unrealistically high number of divisions could wholly account for the amount of acquired skewing observed. Indeed, Abkowitz and coworkers had to decrease the number of stem cells to about 30 before stochastically acquired skewing was evident.³⁰

Discussion

The main finding of this study is a large genetic component in the determination of asymmetric methylation of X-chromosome alleles in the granulocytes of elderly twins. This is in accordance with a recent paper by Christensen and colleagues.²⁹ Our data extend these observations. Not only do we present data comparing X inactivation in granulocytes versus T cells, rather than just peripheral blood, but we also have analyzed the effect in dizygotic as well as monozygotic twins, which permits an estimate of heritability. Furthermore, our estimates of correlation and therefore heritability take into account experimental error. Finally, we have provided a quantitative analysis of acquired stochastic skewing. Our estimate of the heritability of acquired skewing is 68% of the observed variance in granulocytic allele frequencies. Although the conclusion that genetic factors are important is statistically significant, it should be emphasized that the confidence intervals of the estimate of heritability are wide (0%-100%), due to the relatively low number of dizygotic twins analyzed. We believe the most plausible mechanism underlying this phenomenon is genetically determined selection of hematopoietic stem cells (HSCs) with one active X chromosome over those with the other, although other possible mechanisms are discussed later. Stem cells are believed to divide

Table 2. The mean and 95% probability interval for the correlation between twin pairs for T cells and granulocytes

	Monozygotic twin pairs	Dizygotic twin pairs
Granulocytes	0.53 (0.24, 0.75)	0.19 (-0.17, 0.52)
T cells	0.26 (-0.18, 0.64)	0.36 (-0.077, 0.70)

throughout life. After each division, the 2 daughter cells may remain as stem cells or either differentiate or die. The first of these possibilities is believed to occur with a probability of 0.5, so that each stem cell has a probability of 1.0 of continuing as a stem cell after division. Furthermore, good evidence indicates that this choice is stochastically determined, both with respect to fate and time.^{30,40} If the replacement ratio were higher in the stem cells with one X chromosome active than those with the other, then one X chromosome would predominate over the other with passing years. Because this effect would likely be exponential, the absolute difference in ratio from 1.0 need only be relatively small.

To estimate the magnitude of this effect, it is necessary to know the mean rate of stem cell division, which has not been measured directly in humans. However, extrapolating from the rate of telomere loss with age,⁴¹ as well as an independent approach analyzing the randomization of mutant cells with age,³⁹ indicates that human HSCs divide about once per year. It seems unlikely that division occurs much less often. In this case, a difference of 0.02 in the replacement ratio (1.01 for one X chromosome, 0.99 for the other) would lead to a ratio of 3.32 in the ratio of the 2 alleles over 60 years or divisions. Likewise, a difference of 0.002 in the replacement ratio (1.001 for one X chromosome, 0.999 for the other) would lead to a ratio of 1.13. These figures span most of the reported range of acquired skewing. Furthermore, this mechanism, being exponential, is consistent with the degree of acquired skewing being inapparent throughout most of the life span, then apparently increasing markedly over the age of 60. However, longitudinal data from Safari cats imply that the process of genetic selection may be more complex than the simple exponential model considered here.²⁸

The degree of skewing seen in our study was less than that reported by others. We believe the most likely explanation for this discrepancy is that our subjects were slightly younger than the subjects of other studies. Another possible explanation is that both the number of stem cells and their cycling rate increase in old age. Although the number of stem cells is believed to remain approximately constant with respect to age, some evidence indicates that the number may increase in elderly mice.^{42,43}

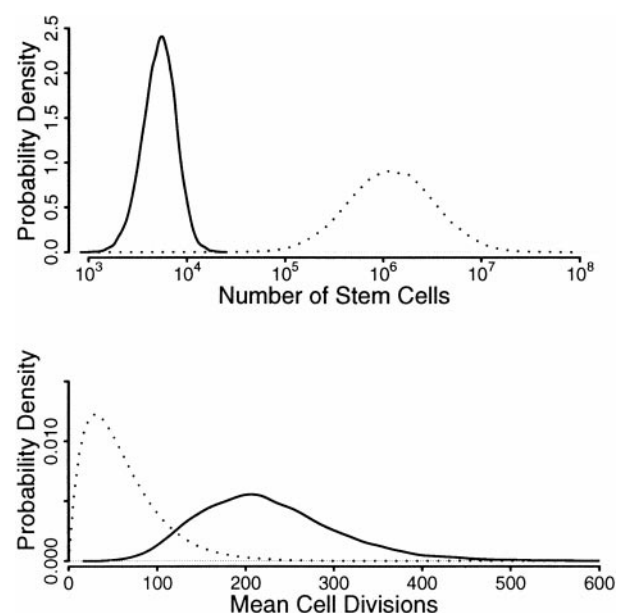


Figure 6. Before and after probabilities from modeling asymmetric stochastic stem cell division. Presented in order to attempt to explain the phenomenon of acquired skewed X inactivation.

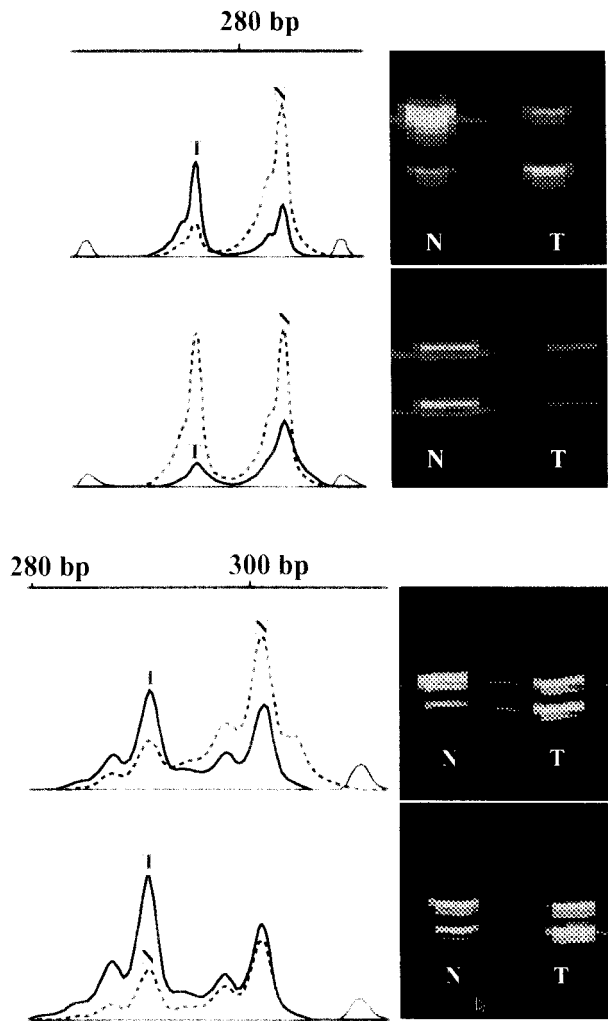


Figure 7. Examples of raw data. On the right hand side of each panel are virtual images of fluorescent bands from the gel. Those from the granulocytes are marked N and those from the T cells are marked T. On the left hand side of the panel are shown the intensities of the corresponding bands on a vertical scale plotted against molecular weight in base pairs (bp) on the horizontal scale. Those from granulocytes are marked in interrupted lines and those from T cells in solid lines. In the lower part of the diagram is shown an example of skewed X inactivation in granulocytes but the drift from that in the T cells is similar in each of the monozygotic twin pairs, which is probably due to genetic selection of one chromosome over the other. In the upper part of the diagram are shown data from one of 2 individuals with skewed X inactivation compared to both their T cells and cognate twin (also shown), which is difficult to explain with a selective mechanism.

Although the degree of acquired skewing in our study is relatively weak, it appears to be common. Perusal of Figures 1 through 4, and the supporting randomization statistic from the first equation, indicate that the deviation of granulocytic from T-cell allele ratios tends to be in the same direction even when the degree of deviation is small. Thus, it appears that most X-chromosome pairs have some degree of differential selective pressure that becomes apparent after many cell divisions. However, the degree of selection varies between different pairs of X chromosomes and so the time of life at which the difference becomes apparent also varies. We would predict that if sufficiently old individuals were studied most would exhibit acquired skewing, even by the rigorous criteria of other authors.^{20,23-25,29}

The second main finding of this study is that there is probably a congenital component in the determination of asymmetric methylation of X-chromosome alleles in the T cells of elderly twins. This may arise from some form of genetic determination or a shared

placental circulation. On one hand, evidence from a study of X-inactivation ratios of monozygotic twins during childhood revealed that the two thirds of monozygotic twins who were monochorionic had more similar allele ratios than those of dichorionic monozygotic twins.⁴⁴ Although the form in which those data were reported precludes exact comparison with our data, when our data for T cells are presented in a similar way the 2 distributions are not easy to distinguish (analysis not shown). This observation suggests that the main reason for the correlation of T-cell ratios in our study is that many twins shared a circulation in utero. On the other hand, the degree of similarity in T cells in our study was similar in both dizygotic and monozygotic twins. In dizygotic twins, many of the alleles could be clearly distinguished from one another and we saw no evidence of a contribution from any "third" alleles. These observations suggest a degree of genetic determination of T-cell allele ratios.

T cells are relatively long lived and so, in studies on X inactivation in adults, they have been considered to reflect X-inactivation status at the time of birth. Furthermore, the status at the time of birth is assumed to be determined randomly, in fact by a binomial process where the number of cells at inactivation is about 7 to 15.¹⁷⁻²⁰ If T-cell inactivation ratios are partly genetically determined, then one or both of these assumptions are wrong. A priori considerations concerning T-cell physiology would suggest the former assumption is inaccurate in at least 2 ways. First, although most T-cell production from HSCs occurs early in life, some production continues throughout life^{45,46} and so skewed HSCs will result in skewed T cells, although the degree of skewing will lag behind that in the granulocytic fraction. Second, T cells comprise a pool of constant size within which clones dynamically expand and contract depending on antigenic exposure.⁴⁷ These processes should also contribute to stochastically determined variance. It is also possible that the latter processes are partly determined by polymorphic X-linked determinants, but this would be expected to be subject to too many uncertainties to make quantitative modeling worthwhile.

Furthermore, the assumption that the distribution of inactivation ratios is determined by a small number of cells at the time of inactivation may also be wrong. Several other explanations are also possible. First, it is possible that some alleles are preferentially amplified in the PCR reaction over others. Although this can be controlled for by comparing prerestriction and postrestriction ratios, our unpublished data (manuscript in preparation) suggest that methylation status may affect the predigestion ratio, making it difficult to quantitate such an effect. Second, the process of X inactivation itself might be subject to polymorphic X-linked determinants.⁴⁸ In particular, it is possible that a residual degree of imprinting at the Xist locus before gastrulation may render the process nonrandom. Our data cannot cast light on these possibilities; longitudinal studies on twinned neonatal blood and other family members would have to be studied. Whatever the mechanism underlying the correlation of T-cell ratios, our data indicate that caution should be exercised in regarding T cells as controls for the original state of X-inactivation.

Although our data indicate that genetic factors are important determinants of acquired skewing, this effect does not account for all of the observed variance. Experimental error accounts for some of the remaining variance, but it seems plausible that asymmetric stochastic cell division accounts for much of the remaining variance. However, in addition to our experimental data, we have provided a quantitative analysis of the likely effects of this mechanism on acquired skewing and have shown that, on current

estimates of stem cell number and turnover, the mechanism is unlikely to account fully for the data. Although it would be possible in principle to use the observed residual variance to estimate stem cell number and turnover, we feel the uncertainties in such a process would make the values of the parameters obtained of little use. It is also possible that environmental effects contribute but these would be difficult to analyze.

Finally, 2 monozygotic individuals exhibited markedly skewed granulocytic X inactivation compared to their respective T cells and compared to their cognate twin (Figures 4 and 7). The degree of discrepancy is beyond that which can be easily explained by any stochastic asymmetric division mechanism superimposed on genetic selection. We suggest that these cases represent overrepresentation of a single clone of cells. It is of interest that this frequency is about the same as the proportion who will go on to develop a hematopoietic stem cell malignancy (myeloproliferative disorders, chronic myeloid leukemia, most acute myeloid leukemia and myelodysplasia in the elderly) over the next 30 years.⁴⁹ However, the peripheral blood count of one of these 2

individuals was not significantly different from that of either their twin or the other individuals in the study. Data on the other individual were not available.

In summary, the mechanisms controlling X-inactivation ratios in normal subjects are complex. Our data show that there is a large genetic component in the determination of this ratio in granulocytes and indicate a smaller genetic component in the ratio of T cells. It seems likely that stochastic asymmetric stem cell division contributes to acquired skewing, but is unlikely to be the major contributor. Two of the 94 informative subjects studied had X-inactivation ratios skewed markedly in the opposite direction to their twin, which may be explained by clones with a proliferative advantage.

Acknowledgments

We thank Rona Morrison for many of the DNA extractions, Gabriella Surdelescu for blood collection, and Kourosh Ahmadi for assistance with the twin analysis.

References

- Lyon MF. Gene inactivation in the X-chromosome of the mouse. *Nature*. 1961;190:372-373.
- Oni SB, Osunkoya BO, Luzzatto L. Evidence for a monoclonal origin of abnormal red cells. *Blood*. 1970;36:145-152.
- Vogelstein B, Fearon ER, Hamilton SR, Feinberg AP. Use of restriction fragment length polymorphisms to determine the clonal origin of human tumours. *Science*. 1985;227:642-645.
- Gale RE, Wainscoat JS. Clonal analysis using X-linked DNA polymorphisms. *Br J Haematol*. 1993;85:2-8.
- Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X-chromosome inactivation. *Am J Hum Genet*. 1992;51:1229-1239.
- Fialkow PJ. Clonal origin of human tumours. *Biochim Biophys Acta*. 1976;458:283-321.
- Wiggans RG, Jacobson RJ, Fialkow PJ, Wooley P, Macdonald JS, Schein PS. Probable clonal origin of acute myeloblastic leukemia following radiation and chemotherapy of colon cancer. *Blood*. 1978;52:659-663.
- Dow LW, Martin P, Moehr J, et al. Evidence for clonal development of childhood acute lymphoblastic leukemia. *Blood*. 1985;66:900-907.
- Gale RE, Bunch C, Moir DJ, Patterson KG, Goldstone AH, Linch DC. Demonstration of developing myelodysplasia/acute myeloid leukaemia in haematologically normal patients after high-dose chemotherapy and autologous bone marrow transplantation using X-chromosome inactivation patterns. *Br J Haematol*. 1996;93:53-58.
- Fialkow PJ, Faguet GB, Jacobson RJ, Vaidya K, Murphy S. Evidence that essential thrombocytopenia is a clonal disorder with origin in a multipotent stem cell. *Blood*. 1981;58:916-919.
- Fialkow P, Gartler SM, Yoshida A. Clonal origin of chronic myelocytic leukemia in man. *Proc Natl Acad Sci U S A*. 1967;58:1468-1471.
- Anger B, Janssen JW, Schrezenmeier H, Hehlmann R, Heimpe H, Bartram CR. Clonal analysis of chronic myeloproliferative disorders using X-linked DNA polymorphisms. *Leukemia*. 1990;4:258-261.
- Brown CJ, Ballabio A, Rupert JL, et al. A gene from the region of the X-inactivation centre is expressed exclusively from the inactive X-chromosome. *Nature*. 1991;349:38-44.
- Brockdorff N, Ashworth A, Kay GF, et al. Conservation of position and exclusive expression of mouse Xist from the inactive X-chromosome. *Nature*. 1991;351:329-331.
- Lovell-Badge R, Avner P. Anti-Xistentialism. *Nat Genet*. 1999;21:343-344.
- Marahrens Y. X-inactivation by chromosomal pairing events. *Genes Dev*. 1999;13:2624-2632.
- Buescher ES, Aling DW, Gallin JI. Use of an X-linked human neutrophil marker to estimate timing of lyonisation and size of the dividing stem cell pool. *J Clin Invest*. 1985;76:1581-1584.
- Puck JM, Stewart CC, Nussbaum RL. Maximum-likelihood analysis of human T-cell X-chromosome inactivation patterns: normal women versus carriers of X-linked severe combined immunodeficiency. *Am J Hum Genet*. 1992;50:742-748.
- Gale RE, Wheadon H, Linch DC. X-chromosome inactivation patterns using HPRT and PGK polymorphisms in haematologically normal and post-chemotherapy females. *Br J Haematol*. 1991;79:193-197.
- Tonon L, Bergamaschi G, Dellavecchia C, et al. Unbalanced X-chromosome inactivation in haematopoietic cells from normal women. *Br J Haematol*. 1998;102:996-1003.
- Tan SS, Williams EA, Tarm PPL. X-chromosome inactivation occurs at different times in different tissues of the post implantation mouse embryo. *Nat Genet*. 1993;3:170-174.
- Fey MF, Peter HJ, Hinds HL, et al. Clonal analysis of human tumours with M27 β , a highly informative polymorphic X-chromosomal probe. *J Clin Invest*. 1992;89:1438-1444.
- Busque L, Mio R, Mattioli J, et al. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood*. 1996;88:59-65.
- Champion KM, Gilbert JG, Asimakopoulos EA, Hinshelwood S, Green AR. Clonal haemopoiesis in normal elderly women: implications for the myeloproliferative disorders and myelodysplastic syndromes. *Br J Haematol*. 1997;97:920-926.
- Gale RE, Fielding AK, Harrison CN, Linch DC. Acquired skewing of X-chromosome inactivation patterns in myeloid cells of the elderly suggests stochastic clonal loss with age. *Br J Haematol*. 1997;98:512-519.
- Migeon BR. Non-random X-chromosome inactivation in mammalian cells. *Cytogenet Cell Genet*. 1998;80:142-148.
- Luzzatto L, Usanga EA, Bienzie U, Esan GF, Fusuan FA. Imbalance in X-chromosome expression: evidence for a human X-linked gene affecting growth of hemopoietic cells. *Science*. 1979;205:1418-1420.
- Abkowitz JL, Taboada M, Shelton GH, Catlin SN, Guttorp P, Kiklevich JV. An X-chromosome gene regulates hematopoietic stem cell kinetics. *Proc Natl Acad Sci U S A*. 1998;95:3862-3866.
- Christensen K, Kristiansen M, Hagen-Larsen H, et al. X-linked genetic factors regulate hematopoietic stem-cell kinetics in females. *Blood*. 2000;95:2449-2451.
- Abkowitz JL, Catlin SN, Guttorp P. Evidence that hemopoiesis may be a stochastic process in vivo. *Nat Med*. 1996;2:190-197.
- Speigelhalter DJ, Thomas A, Best NG, Gilk WR. BUGS: Bayesian Inference Using Gibbs Sampling. MRC Biostatistics Unit, Cambridge. <http://www.mrc-bsu.cam.ac.uk>, 1995.
- Hartl DL, Clark AG. *Principles of Population Genetics*. 2nd ed. Sunderland, MA: Sinauer; 1989.
- Wright, S. *Evolution in natural populations*. *Genetics*. 1931;16:97-159.
- Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics*. 4th ed. Harlow, United Kingdom: Longman; 1996.
- Plomin R, Defries JC, McClearn GE. *Behavioural Genetics, A Primer*. 3rd ed. New York: Freeman; 1998.
- Berardi AC, Wang A, Levine JD, Lopez D, Scadden DT. Functional isolation and characterization of human hematopoietic stem cells. *Science*. 1995;267:104-108.
- Pettengell R, Luft T, Henschler R, et al. Direct comparison by limiting dilution analysis of long-term culture-initiating cells in human bone marrow, umbilical cord blood, and blood stem cells. *Blood*. 1994;84:3653-3659.
- Wang JC, Doedens M, Dick JE. Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay. *Blood*. 1997;89:3919-3924.
- Vickers M, Brown GC, Cologne JB, Kyoizumi S. Modelling haemopoietic stem cell division by analysis of mutant red cells. *Br J Haematol*. 2000;110:54-62.
- Blackett N, Gordon M. "Stochastic"—40 years of use and abuse. *Blood*. 1999;93:3148-3149.
- Ball SE, Gibson FM, Rizzo S, Tooze JA, Marsh JCW, Gordon-Smith EC. Progressive telomere shortening in aplastic anemia. *Blood*. 1998;91:3582-3592.

42. Morrison SJ, Wandycz AM, Akashi K, Globerson A, Weissman IL. The ageing of hemopoietic stem cells. *Nat Med*. 1996;2:1011-1016.
43. de Haan G, Van Zant G. Dynamic changes in mouse hematopoietic stem cell numbers during aging. *Blood*. 1999;93:3294-3301.
44. Monteiro J, Derom C, Vlietinck R, Kohn N, Lesser M, Gregersen PK. Commitment to X inactivation precedes the twinning event in monozygotic MZ twins. *Am J Hum Genet*. 1998;63:339-346.
45. Michie CA, McLean C, Alcock C, Beverley P. Lifespan and human lymphocyte subsets defined by CD45 isoforms. *Nature*. 1992;360:264-265.
46. Richards SJ, Morgan GJ, Hillmen P. Analysis of T-cells in paroxysmal nocturnal hemoglobinuria provides direct evidence that thymic T-cell production declines with age. *Blood*. 1999;94:2790-2799.
47. Green MHL, O'Neill JP, Cole J. Suggestions concerning the relationship between mutant frequency and mutation rate at the hprt locus in human peripheral T-lymphocytes. *Mutat Res*. 1995;334:323-339.
48. Plenge RM, Hendrick BD, Schwartz C, et al. A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. *Nat Genet*. 1997;17:353-356.
49. Cartwright RA, Alexander FE, McKinney PA, Ricketts TJ. *Leukaemia and Lymphoma: An Atlas of Distribution within Areas of England and Wales 1984-1988*. London, United Kingdom: Leukaemia Research Fund; 1990.