X Chromosome Inactivation in the Normal Female Genital Tract: Implications for Identification of Neoplasia

George L. Mutter and Kevin A. Boynton

Department of Pathology, Brigham and Women’s Hospital [G. L. M., K. A. B.], and Harvard Medical School [G. L. M.], Boston, Massachusetts 02115

ABSTRACT

Monoclonal proliferative lesions may be identified by X chromosome inactivation skewing relative to normal polyclonal tissues. We have quantitatively analyzed X-inactivation patterns throughout polyclonal uterine tissues to develop interpretive criteria for recognition of monoclonal neoplasms. Six fresh tissue samples (two samples each of cervix, endometrium, and myometrium) were collected from hysterectomy specimens, and the percentage of androgen receptor (HUMARA) marker allele present on inactive X chromosomes was calculated from a PCR assay. Exact balancing yields 50% of the marker on the inactive X, whereas complete skewing shows either 0 or 100%. X inactivation was similar throughout the tissues of each uterus but was significantly different among the 11 women studied. Comparison of differences in X inactivation between pairs of polyclonal tissue samples within each uterus (XI spread) permitted delineation of cumulative experimental and biological variation of this parameter. Polyclonal-polycional XI spread averaged 10.7 and was independent of the tissue type, sampling site, or the individual studied. Severe baseline skewing of reference polyclonal tissues or contamination of monoclonal tissue by polyclonal cells may reduce the polyclonal-monoclonal XI spread. The extent of X-inactivation skewing necessary to infer a monoclonal process should exceed the 20 or 27 point spread seen, respectively, between 85 and 95% of polyclonal samples.

INTRODUCTION

Tumors in women may be recognized as monoclonal by comparison of X chromosome inactivation patterns in matched neoplastic and normal tissues. This is based on the expected random nature of X inactivation, which occurs at implantation and is conserved in subsequent mitotic cell divisions throughout adulthood (1, 2). Neoplastic proliferations arising from single cells, in contrast, comprise cells that share the same inactivated parental X chromosome; thus, they have an imbalanced or skewed X-inactivation pattern (Fig. 1; Refs. 3–5). Since the introduction of rapid and sensitive PCR-based assays for X-inactivation analysis (6–8), there is renewed interest in utilizing this approach to differentiate neoplastic from nonneoplastic processes. In practice, this involves coanalysis within a single individual of an unknown lesional DNA and a parallel DNA sample from the polyclonal source tissue in which the lesion has arisen. The extent of skewing is compared between the lesonal and normal tissues, and a judgment must be made whether the extent of skewing is different between samples.

Several technical and biological parameters must be controlled to accurately infer clonality status from X-inactivation patterns. PCR bias may be significant, especially with GC-rich repetitive PCR targets, such as the human androgen receptor gene HUMARA (9). In this case, bias may be minimized by frequent use of a high-quality, noncontaminated template DNA or through substitution of 7-deaza-2' -dGTP for dGTP during amplification (9). The latter reduces stability of intramolecular and intermolecular GC base pairing. Quantitative X-inactivation values measured under particular conditions may be validated by demonstration of a linear response with progressive dilutions of monoclonal and polyclonal DNAs (Fig. 2).

Isolated skewing (or imbalance) of X inactivation in a tissue is not a specific indicator of monoclonality because this may be seen in both monoclonal and polyclonal tissues. The population distribution of X-inactivation skewing of polyclonal tissues is a tissue-specific function of the number of component clones and the extent of clonal admixture (5, 10). This was first noticed when phenotypic analysis of polymorphic X-linked protein expression (11) showed substantially skewed X-inactivation patterns (12) in polyclonal human uterine myometrium. Since then, similar skewing in other polyclonal tissues (lymphocytes) has been documented by using a diversity of analytical modalities, including Southern blot analysis (3, 13) and PCR (14).

Partial or complete skewing is an expected finding in tumor DNAs contaminated by polyclonal host cells such as inflammatory, connective, and vascular elements. Such intermediate degrees of skewed X-inactivation within “tumor” samples is determined by both the extent of tumor purity and the baseline level of polyclonal skewing (Fig. 2). In our studies of endometrial precancers, we have often found physically small epithelial lesions to be heavily contaminated by adjacent stroma or inflammatory cells, which are difficult or impossible to completely remove before analysis. Because clonal analysis has great promise in identification of cancer precursors at this and other sites, it has become a pressing matter to define precisely just how much “skewing” is needed to confidently recognize monoclonality.

Systematic delineation of the normal landscape of X inactivation within multiple sites of a control tissue and between tissue types is required to predict how the choice of control polyclonal DNAs may influence clonality interpretation. Ideally, one would like to have a control tissue, the embryonic lineage of which is shared with the lesion, and which contains a homogenous admixture of component multiple clones. We have directly assessed variations in X-inactivation patterns within and among three tissue types of common embryological derivation: uterine cervix, endometrium, and myometrium. This was undertaken in support of our attempts to identify endometrial neoplasms and biologically aggressive endometrial “precancers” by X-inactivation analysis.

MATERIALS AND METHODS

Case Selection. Fresh hysterectomy specimens removed for benign uterine disease (common indications were leiomyomata, prolapse, menometrorrhagia, adenomyosis, pelvic pain, or endometriosis) were obtained from the Women’s and Perinatal Division of the Department of Pathology at Brigham and Women’s Hospital (Boston, MA). Tissue collection for this study was reviewed and approved by the institutional human studies committee. Entry was random and included all specimens available at times of attempted collection. Paired tissue samples measuring approximately 4–9 mm² were excised from the endometrium, cervix, and myometrium (six total) of each hysterectomy specimen, frozen in liquid nitrogen, and stored at −70°C. The two specimens obtained from each tissue came from widely separated sites on opposite sides of the specimen, usually on the right and the left. The genotype of the marker locus HUMARA was screened by PCR amplification of DNA from each hysterectomy by using the primers AR-a/b. Only cases with a heterozygous genotype...
having alleles resolved by a minimum of 4 mm in an 8% polyacrylamide gel were accepted for additional study. After isolation of DNA, each fresh tissue sample was analyzed in duplicate (duplicates had independent digestions, PCR reactions, gels, and quantitation) by using methods described previously (9, 15), which will be summarized in this study.

**DNA Isolation and Digestion.** Frozen tissues were pulverized in liquid nitrogen before lysis by proteinase K and purification by organic extraction and sequential ethanol precipitation by ammonium acetate and sodium acetate, respectively. The final precipitation was followed by careful removal of residual salts through vigorous washing of the pellet in cold 70% ethanol before resuspension in 100 μl of 10 mM Tris (pH 8.0) and 1 mM EDTA buffer.

An undigested aliquot of 2 μg DNA was retained as a control. Two μg of DNA were digested in HhaI buffer with 20 units HhaI overnight, quenched by phenol-chloroform extraction, and ethanol precipitated before PCR with AR-a/b primers.

**PCR Amplification.** Amplification of 200 ng solubilized DNA with primers AR-a and AR-b (15) (5'-CCGAGGAGCTTTCCGAAGATC-3' and 5'-TACGATGGGCTTGGGAGAAA-3', respectively), was performed in a 50 μl PCR reaction mix with 7-deaza-2'-dGTP (9). Trace amounts (final concentration, 50–100 nm) of [³²P]TTP (800 Ci/mmol) were added to the reaction mixture [10 mM Tris (pH 8.4), 50 mM KCl, 20 μg/ml gelatin, 1.5 mM MgCl₂, 0.2 mM dATP, 0.2 mM 7-deaza-2'-dGTP (from Boehringer Mannheim, Indianapolis, IN), 0.2 mM dCTP, and 0.05 mM TTP] to label amplified products. Oligonucleotide primers were added to the reaction mix to 0.3 μM concentration of each. Annealing temperature was empirically optimized at 55°C by comparison of results obtained over a range of temperatures. Thermal cycling in an MJ PTC-100 (MJ Research, Inc., Watertown, MA) thermal cycler followed the sequence: (a) "hot-start" preheating of reaction mixture without Taq polymerase at 97°C for 5 min; (b) 10-min dwell at 85°C, during which 1.25 units Taq polymerase (Perkin Elmer Cetus, Norwalk, CT) were added to each reaction tube; (c) three cycles with prolonged denaturation consisting of 95°C x 4 min, 55°C x 45 s, and 72°C x 90 s; (d) 23 cycles at 95°C x 30 s, 55°C x 45 s, and 72°C x 90 s; and (e) final extension at 72°C for 7 min. PCR products were electrophoresed at 200–500 V in an 8% nondenaturing 0.4-mm thick polyacrylamide gel made in 45 mM Tris-borate-1 mM EDTA. Gels were dried, and autoradiography was performed with preflashed X-ray film with intensifying screens.

**Analysis of PCR Results.** Autoradiogram absorbance was measured with an EC model 910 optical densitometer (EC Apparatus Corp., St. Petersburg, FL), and the resultant plot was integrated by using the GS365W Electrophoresis Data System (Version 2.0; Hoeffer Scientific Instruments, San Francisco, CA). The HMW³ and LMW HUMARA allele PCR products were measured in the paired undigested (U) and HhaI digested (D) lanes. Data in paired lanes were processed further if both lanes had a measurable PCR signal. The X-inactivation ratio was measured as the percentage of the high molecular weight allele present on the inactive X chromosome as:

\[
\text{X inactivation, } \% \text{ HMW} = \left(100 \times \frac{\text{HMW}^0}{\text{HMW}^0 + \text{LMW}^0}\right) \times \frac{\text{LMW}^0}{\text{HMW}^0}
\]

Numerical data were imported into the statistical graphics program Systat (Systat, Inc., Evanston, IL), version 5.0W for analysis and graphing.

**RESULTS**

**Case Selection.** Of the 20 hysterectomy cases accessioned, 3 were rejected because of the homozygosity of the HUMARA locus, and 6
heterozygous cases were rejected because resolution of HUMARA alleles was insufficient for optimal densitometry (electrophoretic resolution, <4 mm). This left 11 hysterectomies for complete evaluation.

**Assay Precision in Fresh and Archival Tissues.** Coefficients of variation of replicate determinations provide an index of assay precision. Of the 66 fresh tissues analyzed, the average coefficient of variation of duplicates was 9.0%, ranging from 0.04 to 50.3%. The coefficients of variation were higher with low X-inactivation values, presumably due to signal measurement inaccuracies that increased with diminishing signal intensity of the digested high molecular weight allele. For example, the highest noted coefficient of variation of 50.3% comes from a mean value of 24%, with a SD of 12.

**X-inactivation Variation within Tissues of Individual Subjects.** Variation of X-inactivation patterns is shown in Fig. 3 between separate samples of each tissue type by individual. A measure of variation within each patient’s tissues is seen graphically by the slope of the connecting lines. The within-tissue differences were not significantly divergent among the three tissues examined (Kruskal-Wallis, P = 0.383).

Each patient had X-inactivation analysis of six independent samples, making it possible to quantitate differences between any two samples from one person as a function of the tissue source. Six samples from each patient were paired in all 15 possible permutations (1 each of cervix-cervix, endometrium-endometrium, and myometrium-myometrium; and 4 each of cervix-endometrium, endometrium-myometrium, and myometrium-cervix), and the “Xi spread” between sample pairs calculated by subtraction. For example, a sample pair with 45.2 and 55.0% X inactivation, respectively, would have an Xi spread of 55 - 45.2, or 9.8. Spread analysis is preferred to SD analysis because (a) difference calculations are more intuitive than are SD calculations; (b) interpretation does not require normalization against the mean; and (c) they are less affected by level-dependent changes in assay precision. There was no significant difference between the Xi spreads obtained for paired samples of the same tissue type compared to those with pair samples of different tissue origin (Kruskal-Wallis, P = 0.599). There were also no differences between sample pair Xi spreads when grouped by side (right/left or anterior/posterior) of origin (Kruskal-Wallis, P = 0.855) or patient of origin (Kruskal-Wallis, P = 0.058). Because we could not discern any specific subset of tissue pairs that had an Xi spread significantly different than another subset, we present an aggregate summary of all 165 tissue pairs in Fig. 4.

**X-inactivation Variation among Individual Subjects.** X-inactivation values for each patient’s cervix, endometrium, and myometrium were calculated by averaging those of the two separate samples. Uterine X-inactivation patterns varied significantly among individuals (ANOVA, P < 0.001), the magnitude of which is shown in Fig. 5.

**Uterine Stem Cell Pool Size.** The mean (ANOVA, P = 0.613) and variance (Levene Test, P = 0.852) of X-inactivation values between tissue types throughout the population of 11 individuals studied were not significantly different. This is consistent with the idea that cervix, endometrium, and myometrium all have the same number of stem cells. The measured population (among individuals) variation in X chromosome inactivation of each tissue was used to calculate the stem cell pool size according to a formula based on a simple binomial distribution: variance observed = pq/N where p and q are the probabilities of each allele occurring.

---

Fig. 3. The X-inactivation ratio is spatially conserved within tissue compartments of individuals. For each patient studied, X-inactivation values from two physically separated samples of each tissue are connected by lines, with the low value on the left axis and the high value on the right axis. Horizontal lines, indicate no difference between samples, with increasing slopes reflecting greater divergence of X-inactivation ratios.

![Fig. 3](image)

Fig. 4. X-inactivation differences between two separate sites sampled from one individual. Frequency distribution of the spread between X-inactivation measurements made between two uterine sites of one individual. Six tissue samples from each of 11 patients were paired in all 15 possible permutations, and the difference in X-inactivation percentages (Xi spread) obtained for each pair. This plot shows proportion of all tissue pairs (ordinate) as a function of Xi spread (abscissa).

![Fig. 4](image)

Fig. 5. Genital tract X-inactivation ratios vary between individuals. The average X-inactivation value for each tissue of 11 individuals studied are plotted. Values for cervix, endometrium, and myometrium are closely clustered for each case but shift significantly (ANOVA, P < 0.001) from one individual to another.

![Fig. 5](image)
abilities of inactivating a particular X (both 0.5), and \(N\) is the stem cell pool size (5, 10). Calculated stem cell numbers for cervix, endometrium, and myometrium were 10.4, 13.6, and 11.8, respectively. Assuming they share a common stem cell pool, we used 12 stem cells to model with a standard binomial theorem the overall expected population distribution of skewed X inactivation. Fourteen \% of all women on this basis should have uterine skewing of less than or equal to 25\% or greater than or equal to 75\%.

**DISCUSSION**

**X Inactivation Is Homogenous within Uterine Tissues of Common Embryonic Lineage.** For each individual studied, X-inactivation patterns were conserved within and among samples of the cervix, endometrium, and myometrium. Variation between tissue locations of each individual (overall coefficient of variation of two separate samples of each tissue \(= 16.6\%\); Fig. 3) was only slightly greater than analytical variation in repeat analyses of one DNA sample (coefficient of variation \(= 9\%\)). Although this difference is statistically significant (ANOVA, \(P = 0.001\)), the majority of variation seen among tissue samples can be attributed to analytical rather than biological variation. We conclude that at the current level of analytical precision, a single pool size \(5, 10\). Calculated stem cell numbers for cervix, endometrium, and myometrium were 10.4, 13.6, and 11.8, respectively. Assuming they share a common stem cell pool, we used 12 stem cells to model with a standard binomial theorem the overall expected population distribution of skewed X inactivation. Fourteen \% of all women on this basis should have uterine skewing of less than or equal to 25\% or greater than or equal to 75\%.

**Criteria to Identify Uterine Monoclonal Proliferations.** Matched normal polyclonal tissues must be analyzed in parallel with lesional samples for accurate interpretation of X-inactivation data. Our data suggest that endometrium, myometrium, and cervix may be interchanged as controls where needed. The size of the control and lesional tissue samples must be equivalent to avoid differential confounding effects of clonal patch size as discussed below.

A 20 percentage-point shift of X-inactivation percentage between control polyclonal and unknown lesional tissues is the minimal amount required to identify a monoclonal population in the lesional sample. This is based on the observation that shifts beyond this magnitude are encountered between only 15\% of normal uterine polyclonal tissue pairs (Fig. 4). As the Xi spread between normal control and lesional tissue rises, so does the level of confidence that the lesional tissue is not likely to be polyclonal. For example, 90\% of normal uterine tissues have a Xi spread \(\leq 23\), and 95\% have a spread \(\leq 27\). These differences between polyclonal tissues occur because of the combined effects of analytical error and biological variation, and the guidelines stated here assume experimental error is minimized by performing all assays in duplicate. We have not attempted to identify specific sources of analytical error in these experiments beyond our prior analysis of PCR amplification kinetics with primers AR-a/b (9), but theoretical possibilities include a slightly nonlinear PCR reaction, variable completeness of \(HhaI\) digestion, and imprecision in the measurement of PCR products. Linear amplification during PCR is confirmed in Fig. 2 and enhanced by the use of 7-deaza-2'-dGTP (9) in these experiments. Quantitation of PCR products might be more precise by using a system with greater dynamic range than that of X-ray film, such as scintillation counting of excised bands or use of an automated phosphor imaging device.

Sensitivity of detection of monoclonal proliferations is a combined function of the magnitude of polyclonal background skewing, the relative direction of background and monoclonal skewing, and the extent of contamination by polyclonal cells. Interaction of these factors in defining the Xi spread between reference and lesional DNA is seen in Fig. 2. In patient X, monoclonal tissue contaminated with 25\% polyclonal DNA generates a 20 percentage-point shift between control and lesional samples. In contrast, baseline polyclonal skewing in patient Y is so extreme that even pure tumor is only deviated 15 percentage points from the polyclonal level. Calculations of uterine stem cell numbers estimate that approximately 14\% of all women will have polyclonal skewing favoring one allele at a level of \(\leq 25\%\) or \(\geq 75\%\) X inactivation.

**Nonquantitative Visual Assessment of Allelic Skewing in Autoradiograms.** Direct visual assessment of PCR product relative autoradiographic intensity is necessarily less precise than quantitative densitometry, but may suffice for interpretation of clonality results. Typically four lanes must be compared; undigested and digested DNAs from normal and lesional tissue. The purpose of undigested lanes is to assess labeling and amplification efficiency of PCR products, usually constant features within a case. If both alleles are approximately equally represented in these lanes, one may proceed to assess the distribution of inactive X-linked alleles by direct comparison of normal and lesional DNAs that had been predigested with \(HhaI\). Visual analysis tends to focus on band-relative intensity within a lane, rather than percentage distributions of all signal. We have visually examined band patterns of DNAs with allelic shifts quantitated previously by densitometry (such as the titration of Fig. 2), and we conclude that we can recognize shifts in intensity caused by Xi spreads of as little as 15–20%, a level at or below that which specifically indicates a monoclonal growth pattern. Additional gradation of levels of skewing may be acquired with experience and permit discrimination between minor variations seen within polyclonal tissues and larger Xi spreads that are more likely associated with monoclonal processes.

**Tissue-specific Determinants of X-inactivation Patterns.** Tissue-specific patterns of X inactivation may be attributed to the stem cell composition of that tissue, the physical geometry of clonal admixture, and selective pressures for clonal expansion or contraction. Conservation of X inactivation throughout the uterus implies that these factors are constant within this organ, although this may not be the case with nonuterine tissues. Furthermore, we cannot exclude the possibility that infrequent events may occasionally create greater variation than what is represented in our sample. Variables that relate to establishment of tissue-specific differences in X inactivation will be discussed briefly.

Homogeneity of X inactivation among different tissues of common embryological lineage is expected where there is inclusive and unbiased clonal propagation during entry into divergent differentiation pathways. On these grounds it is likely that other Müllerian structures, such as the fallopian tube, will mirror patterns of X inactivation seen in the uterine corpus itself. Descriptive embryological studies may be a better basis to infer proximity of stem cell relatedness of different tissues than spatial or morphological proximity in adult tissues. For example, the squamous epithelia of the cervix and vagina are created in part (vagina and exocervix) by incursion of perineal epidermis that traverses the urogenital sinus to meet the Müllerian tube, and in part (transformation zone) by squamous metaplasia of the Müllerian endocervix. Our deep tissue samples from the cervix, which contained fibromuscular stroma and endocervical glands, are embryologically closer to the squamocolumnar junction (transformation zone) than to the squamous exocervix.

Even if the initial clonal allocation into derivative tissues is symmetrical and equivalent, mutations of genes subject to X inactivation can produce global or tissue-specific expansion of clones either with inactive (16) or active (17) mutant alleles. Female carriers of X-linked mental retardation with \(\alpha\)-thalassemia, a disease caused by the mutation of the \(A7R\)-X gene (16), are known to have selective survival in all somatic tissues examined of clones which have inactivated the mutant allele.

Furthermore, tissue-specific requirements for a functional X-linked gene in carrier mutant heterozygotes can dramatically alter the normal
pattern of conservation of X inactivation among embryologically related tissues. In female carriers of X-linked severe combined immunodeficiency, there is heavily skewed (nonrandom) X inactivation in B and T lymphocytes but balanced inactivation in neutrophils (18). Presumably, preferential survival occurs in lymphocytes because they require a normally functioning gene, whereas neutrophils are spared clonal propagation bias because they do not need products from the affected locus.

Physical mixing of component clones within an embryonic or adult tissue may vary, creating tissue specific heterogeneity of X inactivation. If a progenitor tissue contains an asymmetrical distribution of clones (incomplete mixing during embryogenesis), derivative tissues created by physical partitioning and differentiation will have divergent clonal compositions. In adult polyclonal tissues, physical distribution of multiple clones can be visualized as a confluent series of "patches," a volume occupied by contiguous cells from the same clone. Soon after completion of cell division, progeny cells remain in close proximity to the parent cell, becoming admixed with other clones through cellular movement or tissue remodeling. If the goal is to obtain a sample representative of the overall clonal composition of a tissue, then sample volume must exceed patch size. Patch size throughout the uterus must be less than the 4–9-mm³ volume of samples used in these experiments because we never saw complete skewing indicative of sampling only a single clone. Others (12) have estimated maximum myometrium patch size at 1 mm³, suggesting a possible range of 1–10,000 cells.

Detection of monoclonal lesions using the described PCR assay for X inactivation has a maximum theoretical sensitivity of 75%, based on comparison of uncontaminated monolocal to polyclonal DNAs in a population where 10% of individuals will have noninformative homozygous HUMARA genotypes, and an additional 14–15% will demonstrate confounding X-inactivation skewing of polyclonal tissues. Conservation of X inactivation throughout the uterine cervix, endometrium, and myometrium confirms that they are interchangeable as polyclonal controls for these assays. In practice, sensitivity is reduced further by technical limitations in resolution of similar molecular weight heterozygous alleles and contamination of monolocal tumors with polyclonal cells. The spectrum of X-inactivation spread between two control polyclonal tissues was used to develop a probability profile of expected Xi spreads between polyclonal samples. As the spread between control and unknown samples exceeds that expected between polyclonal tissues, the unknown sample may be identified as monoclonal with increasing specificity.

REFERENCES
X Chromosome Inactivation in the Normal Female Genital Tract: Implications for Identification of Neoplasia

George L. Mutter and Kevin A. Boynton


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/55/21/5080

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.