An Integrative Approach to Assess X-Chromosome Inactivation Using Allele-Specific Expression with Applications to Epithelial Ovarian Cancer

Nicholas B. Larson[1], **Zachary C. Fogarty**[1], **Melissa C. Larson**[1], **Kimberly R. Kalli**[2], **Kate Lawrenson**[3,4], **Simon Gayther**[4], **Brooke L. Fridley**[5], **Ellen L. Goode**[6], and **Stacey J. Winham**[1,§]

[1]Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

[2]Department of Medical Oncology, Mayo Clinic, Rochester, MN, USA

[3]Women's Cancer Program, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[4]Center for Bioinformatics and Functional Genomics, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[5]Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, Tampa, FL, USA

[6]Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

Abstract

X-chromosome inactivation (XCI) epigenetically silences transcription of an X chromosome in females; patterns of XCI are thought to be aberrant in women's cancers, but are understudied due to statistical challenges. We develop a two-stage statistical framework to assess skewed XCI and evaluate gene-level patterns of XCI for an individual sample by integration of RNA sequence, copy number alteration, and genotype data. Our method relies on allele-specific expression (ASE) to directly measure XCI and does not rely on male samples or paired normal tissue for comparison. We model ASE using a two-component mixture of beta distributions, allowing estimation for a given sample of the degree of skewness (based on a composite likelihood ratio test) and the posterior probability that a given gene escapes XCI (using a Bayesian beta-binomial

§CORRESPONDING AUTHOR: Stacey J. Winham, Ph.D., Mayo Clinic, 200 First Street SW, Rochester, MN 55905; (507) 266-3662 (phone), (507) 284-1516 (fax); winham.stacey@mayo.edu.

CONFLICT OF INTEREST: The authors declare no potential conflicts of interest.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Mayo Clinic Institutional Review Board and all patients provided informed consent.

AVAILABILITY OF DATA AND MATERIAL The datasets used during the current study are available from the authors upon reasonable request.

COMPETING INTERESTS The authors do not declare any competing interests.

mixture model). To illustrate the utility of our approach, we applied these methods to data from tumors of ovarian cancer patients. Among 99 patients, 45 tumors were informative for analysis and showed evidence of XCI skewed towards a particular parental chromosome. For 397 X-linked genes, we observed tumor XCI patterns largely consistent with previously identified consensus states based on multiple normal tissue types. However, 37 genes differed in XCI state between ovarian tumors and the consensus state; 17 genes aberrantly escaped XCI in ovarian tumors (including many oncogenes), whereas 20 genes were unexpectedly inactivated in ovarian tumors (including many tumor suppressor genes). These results provide evidence of the importance of XCI in ovarian cancer and demonstrate the utility of our two-stage analysis.

Keywords

Bayesian; mixture model; ovarian cancer; RNA-Seq

INTRODUCTION

In females, X chromosome inactivation (XCI) is a developmentally regulated process that randomly silences transcription of one of the two homologous copies of the X chromosome, resulting in equivalent gene expression dosage with males[Ross, et al. 2005]. Unlike the active copy of the X chromosome (X_a), the inactive X (X_i) forms a compact heterochromatin structure known as the Barr body[Ross, et al. 2005]. This process is epigenetic in nature, dictated by up-regulation of the non-coding RNA encoded by *XIST* and resulting in *cis*-acting transcriptional silencing of the chromosome by DNA hypermethylation and enrichment of histone modifications. However, an estimated 9%–14% of X-linked genes escape XCI and are transcribed on both chromosomes, even beyond the pseudo-autosomal regions[Cotton, et al. 2013]. Furthermore, X-linked genes can demonstrate variability in escape status across tissue types, as well as variability across individuals at the gene-level[Cotton, et al. 2015]. Consequently, XCI may contribute to sexual dimorphism and female phenotypic diversity through the female-specific epigenetic regulation of X gene expression.

The plasticity of XCI may also play a role in disease, particularly women's cancers. Although acquired genomic alterations are often of primary interest in tumorigenesis, recent studies have demonstrated that epigenetic changes also occur in the tumor environment, as reviewed in [Sharma, Kelly, & Jones 2010]. Relevant to XCI, Barr body loss has been observed in ovarian and breast cancer tumors[Barr & Moore 1957; Pageau, Hall, Ganesan, Livingston, & Lawrence 2007]. Although the phenomenon of Barr body loss has largely been attributed to structural alteration via concomitant deletion of the X_i and duplication of X_a, recent breast cell cancer line studies have indicated the epigenetic erosion of XCI and local reactivation of genetic transcription on the X_i[Chaligne, et al. 2015]. Aberrant de-activation of X-linked tumor suppressors and/or re-activation of oncogenes may result in substantial dysregulation of cancer-related gene expression and contribute to tumor progression. In fact, it is thought that tumor suppressor genes that escape from XCI may contribute to the sex bias observed for some cancers[Dunford, et al. 2016]. Moreover, the X chromosome is enriched with hormone-related genes, and disruption of XCI could play a

role in hormonal carcinogenesis[Henderson & Feigelson 2000]. However, practical analysis of XCI is complicated by the complex biology, cell and tissue-specificity, and difficulty in obtaining direct measurements, and patterns of potential XCI aberrations in women's cancers are currently understudied[Chaligne & Heard 2014].

While studying XCI in model organisms is possible via marking maternal and paternal chromosomes[Berletch, Yang, Xu, Carrel, & Disteche 2011], study in humans is more complicated. Outside of cancer, previous methods to study XCI at the gene level in humans have utilized rodent/human cell hybrids, total expression or allelic expression data derived from a microarray, or DNA methylation data[Carrel & Willard 2005; Cotton, et al. 2013; Cotton, et al. 2015]. Single cell sequencing studies have also begun to emerge[Tukiainen, et al. 2016]. These previous approaches have several limitations; for example, reliance on cell hybrids has limited clinical utility. Methods relying on methylation are problematic because it is an indirect method to measure XCI; DNA methylation is one epigenetic mechanism by which *XIST* silences expression of the X_i, but other epigenetic factors are also involved. The use of microarray-derived expression data represents a more direct approach to measure the genes that are expressed on the X_a; however, these data are subject to potential confounding due to batch effects. Previous studies have also utilized male samples as a comparison population[Cotton, et al. 2015; Sharp, et al. 2011]; however, female-specific diseases such as ovarian cancer necessitate analytical methods that are independent of male samples. Additionally, previous approaches do not provide an estimated probability of escape for a given gene for a given individual. Furthermore, previous methods cannot distinguish XCI from possible copy number deletions or loss of heterozygosity (LOH), where only a single copy of the X chromosome remains. Lastly, previous analyses have focused on a single genomic data type, whereas the process of XCI is inherently multi-omic, involving genetic, epigenetic, and expression components.

The quantification of allele-specific expression (ASE), the relative proportion of transcribed mRNA attributable to each homologous copy of each gene, has successfully been utilized in association studies by correlating proximal genetic variation with allelic imbalance[Sun 2012]. Similar principles used to study these *cis*-acting expression quantitative trait loci on the autosomes can be extended to characterize gene-level patterns of XCI. For a given cell, mono-allelic expression suggests that the gene undergoes XCI and is not transcribed on the X_i; in contrast, evidence of balanced allelic expression of an X-linked gene indicates that the gene escapes XCI. In a tissue sample consisting of multiple cells, however, the ability to make this distinction is complicated by the potential for cellular heterogeneity. ASE from a tissue sample where the maternal chromosome is inactivated in some cells, but the paternal chromosome is inactivated in others will not be informative for gene-level XCI, as all genes will exhibit relatively balanced allelic expression regardless of XCI status. In contrast, if the X_i is heavily skewed toward a specific parental chromosome, discrimination between active and inactivated genes may be possible via patterns of ASE imbalance. In addition to XCI, ASE imbalance of X-linked genes may also manifest due to overlapping copy number alterations (CNAs); for instance, copy-neutral LOH would lead to only a single distinct allele being expressed within the impacted segment. Furthermore, an approach to study XCI in the tumor needs to be informed by expected XCI patterns in normal tissue. Consequently,

chromosome-wide studies of XCI require advanced statistical methods that take these properties into consideration.

To address these concerns, we have developed a novel integrative approach that utilizes allele-specific expression data to directly measure gene-level patterns of XCI for a given tumor sample, relative to expected patterns based on normal tissue, and does not require analogous male samples for comparison (Figure 1). Particularly, in this study, we develop the statistical framework for a two-stage approach for the evaluation of gene-level patterns of XCI that relies on the integration of tumor RNA sequence, CNA, and germline genotype data and apply these methods to study XCI tumor escape patterns in a collection of ovarian cancers. First, we leverage ASE to identify tumors with evidence of the X_i heavily skewed toward one parental chromosome to address potential confounding with XCI heterogeneity. We then estimate the tumor XCI escape status for each gene using baseline patterns informed by consensus states from multiple tissue types. Finally, we discuss our findings and highlight future research directions for investigating XCI in cancer using multi-omic applications.

METHODS

Study Participants

Study subjects included a total of N = 99 patients from the Mayo Clinic enrolled into an IRB-approved protocol within one year of diagnosis of pathologically confirmed primary invasive epithelial ovarian, fallopian tube, or primary peritoneal cancer between 2000 and 2009 (Table S1). All tissues were snap frozen immediately following surgery and stored at −80°C as a source of RNA; a gynecologic oncologist reviewed each sample to confirm diagnosis and ensure at least 70% tumor content in tumor samples and the absence of tumor in normal samples. Peripheral blood drawn prior to chemotherapy was used as a source of germline DNA.

Tissue RNA Sequencing

RNA was sequenced on an Illumina HiSeq2000 with six samples per lane in two batches. For 15 samples, 500ng RNA was used to generate polyA libraries using the Illumina TruSeq kit, and samples were run with paired end 50bp reads. The remaining 84 samples had 1μg RNA treated with riboZero and libraries were made using the Illumina TruSeq Stranded Total RNA kit, and samples run with 100bp paired end reads.

Genotyping

Germline DNA was genotyped with Illumina Infinium Beadchips (for all cases, OncoArray-500k; additionally for 81 cases, Human610-Quad; for 18 cases, Omni2.5–8), and a subset of 80 participants were also genotyped using a customized Affymetrix array targeting rare variants[Winham, et al. 2016]. Genotypes were merged at each SNP using the most comprehensive array for variants assayed on multiple platforms. In addition, tumor DNA was genotyped using the Infinium BeadChip (OncoArray-500k) and was utilized for CNA analysis (described below).

Copy Number Aberration Segmentation Analysis

To address potential confounding via structural alterations, we applied OncoSNP[Yau, et al. 2010] to call large segmental (low resolution) tumor CNAs using the OncoArray 500k data. We utilized the paired 'tumor-normal' mode, 'intratumor' and 'stromal' options, and assumed normal contamination levels of 0%–30%.

Quantifying Allele-Specific Expression

RNA sequence reads are ASE-informative only if they overlap one or more expressed heterozygous single nucleotide polymorphisms (eSNPs). First, germline genotypes were chromosomally pre-phased using SHAPEIT2 and imputed using IMPUTE2 with the 1000 Genomes as a background set, resulting in two phased chromosome-wide haplotypes denoted (H_1, H_2). Imputed genotypes were retained based upon an allele dosage $r^2 > 0.25$, minor allele frequency (MAF) 1%, and genotype posterior probability > 0.5. To quantify ASE per sample, haplotype-specific BAM files (one each for H_1, H_2) of the aligned RNA sequence reads were produced for each sample using the *asSeq*[Sun 2012] package in R 3.0.2, and gene-level ASE read counts were generated using HTSeq version 0.5.3p9 (mode = intersection-nonempty). ASE data for genes that at least partially overlapped with genomic segments exhibiting structural alterations (duplications, deletions, and/or loss of heterozygosity) for a given sample were excluded. Batch effects were assessed by comparing per-sample median ASE across genotyping and RNA sequencing batches.

Statistical Model Assumptions

For a given sample with observed ASE data for G total X-linked genes, let (Y_{g1}, Y_{g2}) represent the bivariate vector of ASE reads mapped to the two phased X chromosome haplotypes (H_1, H_2) for gene $g \in \{1, …, G\}$ and define $N_g = Y_{g1} + Y_{g2}$. Let G_E and G_I respectively denote the sample-specific subsets of genes that are transcriptionally active (i.e., escaping) and inactive on the X_i. We assume Y_{g1} follows a binomial distribution with size N_g and success rate p_g, such that

$$\Pr(Y_{g1}=y; N_g, p_g) = \binom{N_g}{y} (p_g)^y (1-p_g)^{N_g-y} \tag{1}$$

where we assume $p_g = 0.5$ for all $g \in G_E$ regardless of XCI skewness, while p_g 0.5 for $g \in G_I$ when the sample corresponds to skewed XCI. Due to random phase switch errors from statistical phasing of genotypes[Delaneau, Zagury, & Marchini 2013], we assume the Y_{g1} reads mapped to H_1 are overall equally likely to correspond to either parental chromosome across genes; we also assume Y_{g1} is subject to extra-binomial variation[Pickrell, et al. 2010]. To accommodate over-dispersion and unknown X homolog membership, we assume p_g follows a mixture of beta distributions (Figure 2), such that

$$f_{p_g}(p; a_I, b_I, a_E) = \begin{cases} \frac{1}{2}\left(Beta(a_I, b_I) + Beta(b_I, a_I)\right) & \text{if } g \in \mathscr{G}_I \\ \frac{1}{2}\left(Beta(a_E, a_E) + Beta(a_E, a_E)\right) & \text{if } g \in \mathscr{G}_E \end{cases} \quad (2)$$

where *Beta*(*a*, *b*) is defined as $f(x; a, b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$ and B(*a*, *b*) is the beta function. Note that $f_{p_g}(p|g \in G_E)$ reduces to *Beta*(a_E, a_E), unimodal and symmetric about 0.5. For modeling purposes, we define $Y_g = \min(Y_{g1}, Y_{g2})$ as the ASE of the least expressed allele and assume Y_g follows a folded distribution[Porzio & Ragozini 2009]. Conditional on XCI gene set membership, the probability $Y_g = y$ for $y \leq \lfloor \frac{N_g}{2} \rfloor$ can be defined as

$$\Pr(Y_g = y; N_g, (a,b)) = \binom{N_g}{y}\left(1 - \frac{\delta(y, N_g)}{2}\right)(B(y+a, N_g - y + b) + B(y+b, N_g - y + a)) \quad (3)$$

where $\delta(y, N_g) = 1$ if $y = N_g - y$ and 0 otherwise, and (*a*, *b*) is equivalent to (a_I, b_I) or (a_E, a_E) for G_I and G_E, respectively. To simplify prior specification and improve interpretation, we adopt a mode-dispersion parameterization (*m*, θ) of the beta distribution *Beta*(*a*, *b*), with mode $m = \frac{a-1}{a+b-2}$ and dispersion parameter $\theta = a + b$, for $m \in [0, 0.5]$ and $\theta \geq 2$. Mode parameter $m_E = \frac{a_E - 1}{2a_E - 2} = 0.5$ simplifies to a fixed constant for genes in G_E while $m_I \leq 0.5$ indicates skewed XCI. Note that m_I can be interpreted as a measure of sample skewness.

Consensus XCI Status Across Multiple Tissues

To leverage prior knowledge of genes that escape XCI, we applied the results of Balaton et al.[Balaton, Cotton, & Brown 2015]. In brief, the authors combined analysis results across three prior studies [Carrel & Willard 2005; Cotton, et al. 2013; Cotton, et al. 2015] that examined XCI across 27 tissue types using multiple genomic approaches (cell hybrids, gene expression, DNA methylation). Based on the patterns of XCI across these three studies, genes were categorized into eight possible consensus states: escape, mostly escape, variable escape, mostly variable escape, subject, mostly subject, discordant, and no call (Table S2). As these studies were based on tissue types available in both sexes, XCI patterns for female-specific tissues, including ovarian epithelium, were not available.

Stage 1: XCI Skewness Testing

ASE from a tissue sample will not be informative for gene-level XCI if the X_a is sufficiently heterogeneous in representation by maternal and paternal chromosomes, since all genes will exhibit relatively balanced allelic expression ratios (i.e., $0.5 = m_E \approx m_I$). In contrast, if the X_a is heavily skewed toward a specific parental chromosome, discrimination between escape and inactive genes may be possible via downstream model-based clustering. To identify samples demonstrating evidence of skewed XCI, we evaluate allelic expression ratio imbalances of the subset of genes *a priori* likely to be inactive on the X_i based upon consensus XCI gene status calls from Balaton et al.[Balaton, Cotton, & Brown 2015], denoted $\mathscr{G}_I^*$, which encompass genes declared to be "subject" or "mostly subject" to XCI.

We similarly define subset $\mathscr{G}_E^*$ for genes declared to be "escaping" or "mostly escaping" XCI. All remaining XCI consensus states (i.e., discordant, variable escape, mostly variable escape, no call) are treated equally as unknown.

Under the null condition of random (non-skewed) XCI, the distribution of ASE reads for genes $g \in \mathscr{G}_I^*$ is comparable to genes $g \in \mathscr{G}_E^*$, equivalent to testing H_0: $m_I = 0.5$ versus H_1 : $m_I \neq 0.5$. For each sample, we define $\boldsymbol{Y}$ to be the vector of observed Y_g for genes $g \in \mathscr{G}_I^*$ and derive the maximum likelihood estimates $(\hat{m_I}, \hat{\theta_I})$ using box-constrained numerical optimization of the log-likelihood function, $\ell(m_I, \theta_I; \boldsymbol{Y}) = \Sigma_g \ln(\Pr(Y_g; N_g, m_I, \theta_I))$ under the null and alternative hypotheses, with $\Pr(Y_g) = 1$ for genes corresponding to $N_g = 0$. We then compute a likelihood ratio test (LRT) statistic $\Lambda = -2(\ell(\hat{m_I}, \hat{\theta_I}; \boldsymbol{Y}) - \ell(0.5, \hat{\theta_I}; \boldsymbol{Y}))$. Since the null condition $m_I = 0.5$ is on the boundary of the support for m_I, we estimate the null distribution of Λ by computing Λ_b for $b = 1, \ldots, B$ parametric bootstrap samples under the fitted null model[Mclachlan 1987]. The LRT P-value is computed as $P = \frac{1 + \sum_{b=1}^{B} I(\Lambda_b \geq \Lambda)}{B+1}$.

Stage 2: XCI Escape Gene Analysis

Conditional on sample XCI skewness, we expect the two mixture components comprising $f_{p_g}(p)$ to be identifiable, corresponding to the latent class categories of 'inactive' or 'escape'. We opt for a Bayesian approach to inference and fit the model using Markov Chain Monte Carlo (MCMC) methods. We assume a flat *Uniform*(0,0.5) prior distribution on m_I. We additionally assume $\theta_I \neq \theta_E$ by modeling $\theta_I = \theta_E + \theta_{jI}^*$, and set priors $\pi(\theta_I^*)$ and $\pi(\theta_I)$ to vague Gamma distributions. Finally, we left-truncate $\pi(\theta_E)$ at 3.0 to ensure unimodal distributions for p_g.

Absent of any prior knowledge, we may define flat independent prior probabilities of a gene $g \in G_E$, η_g, such that $\eta_g = 0.5$. Alternatively, we may leverage the consensus XCI states for the genes of interest to assign gene-specific informative priors. For our purposes, we assigned 'subject', 'mostly subject', 'mostly escape', and 'escape' statuses η_g values of 0.2, 0.3, 0.7, and 0.8, respectively. All other consensus states were classified collectively as unknown and assigned a prior of $\eta_g = 0.5$. Under either prior definition, our measure of interest is the posterior probability of escape for gene g and sample j, $PP_{jg} = P(g \in G_{jE}|Y_j)$:

$$PP_{jg} = \frac{\eta_g \cdot \Pr\left(Y_{jg} = y; N_{jg}, (0.5, \theta_E)\right)}{\eta_g \cdot \Pr\left(Y_{jg} = y; N_{jg}, (0.5, \theta_E)\right) + (1 - \eta_g)\Pr\left(Y_{jg} = y; N_{jg}, (m_{jI}, \theta_{jI})\right)} \quad (4)$$

Data Simulations - Skewness

To generate realistic XCI ASE data, we based our simulations on autosomal ASE data from chromosome 22, derived from available normal tissue samples. These data provide biological distributions of allelic expression for genes not undergoing epigenetic silencing due to XCI (i.e. the null distribution of balanced allelic expression) and allow us to minimize distributional assumptions. Briefly, consider an autosomal gene with observed values of (Y_{g1}, Y_{g2}), (y_{g1}, y_{g2}), and $n_g = y_{g1} + y_{g2}$. To simulate data from a hypothetical gene $g \in G_E$

with total ASE of N_g^{sim}, we sampled N_g^{sim} reads from H_1 and H_2 at the corresponding proportions y_{g1}/n_g and y_{g2}/n_g. For $g \in G_I$ with some true skewness level m_I, these proportions were respectively multiplied by $\frac{1}{m_I}$ and $\frac{1}{1-m_I}$ prior to sampling. Data were generated such that an observation (y_{g1}, y_{g2}) was randomly selected across all normal samples and genes that satisfied $n_g \geq 100$ for each simulated expression result.

To evaluate Type I error and power for the skewness test under a variety of conditions, we considered skewness values of $m \in (0.1, 0.2, 0.3, 0.4, 0.5)$, ASE expression levels of $N_g \in (10, 20, 30)$, and total number of genes evaluated $L \in (10, 20, 30)$, such that $m_I = 0.5$ defines the null condition of random XCI. For each unique simulation condition, a total of 1000 replications were run, with all testing results declared significant at an α-level of 0.05. The number of parametric bootstrap samples B was fixed at 500 in all settings. Power for our skewness test is also dependent upon the accuracy of $\mathscr{G}_I^*$ when selecting a subset of genes for testing. In practice, G_I and $\mathscr{G}_I^*$ may not be equivalent for a variety of reasons, including underlying XCI differences by tissue type as well as discordant XCI gene activity within a given sample. To additionally evaluate robustness to the accidental inclusion of escape genes in the skewness test due to sample-specific differences between $\mathscr{G}_I^*$ and G_I, we considered contamination of 20% of the tested genes to be simulated as truly in G_E. We evaluated power under the same conditions above for $m_I \in (0.1, 0.2, 0.3, 0.4)$.

Data Simulations – Escape Gene Analysis

The same simulation principles as described above for skewness testing were applied to generate simulated XCI data for XCI escape gene analysis. To generate ASE profiles for analysis, we simulated data under conditions $N_g \in (10, 20, 30)$ and $m_I \in (0.1, 0.2, 0.3)$. For a given sample, ASE data were generated for $G = 100$ hypothetical genes with a fixed XCI escape gene proportion of 15% (i.e., 15 genes in G_E). All analyses were conducted using the Bayesian software Stan[Carpenter, et al. 2015] via the *rstan* package and the statistical software R 3.3.1, and convergence was evaluated using standard diagnostics and visual inspection of posterior samples from two independent chains of 2500 iterations after a burn-in of 2500. PP_g for each eligible gene was calculated based on equation (4) using Monte Carlo integration over the combined joint posterior sample of the model parameters.

To investigate the influence of informative priors, we considered two scenarios: lack and presence of discordant genes (i.e. genes inconsistent with their assumed prior state). For the former, we set the XCI consensus state for the genes in G_E to be five each for 'escape', 'mostly escape' and 'unknown'. For the remaining 85 genes in G_I, 30 were designated as 'subject', 15 as 'mostly subject', and 40 as 'unknown'. These proportions approximate those observed in the XCI consensus state data. To simulate the presence of discordant genes, we selected 5 'subject' or 'mostly subject' genes to be truly in G_E, and 5 'escape' or 'mostly escape' genes to be truly in G_I. Performance was evaluated based on the ability of PP_g to discriminate between true XCI-E and XCI-S genes, estimated by classification accuracy, sensitivity, and specificity of escape gene identification using a threshold of $PP_{jg} > 0.5$.

Extensions to Multi-sample Analysis

In practice, many genes may have low expression for a given sample, and the expected number of informative XCI escape genes per sample could be small. To improve inference, we extend the defined model above in (1) and (2) to a multi-sample approach to borrow information across samples and genes, where G_{jI} and G_{jE} respectively represent sample-specific active and inactive gene subsets for sample $j \in 1, \dots J$. To accommodate differences in XCI skewness across samples, we define

$$f_{p_{jg}}(p|g \in \mathcal{G}_{jI}) = \tfrac{1}{2}\left(Beta(m_{jI}, \theta_{jI}) + Beta(1 - m_{jI}, \theta_{jI})\right) \quad (5)$$

For XCI escape genes, we assume $\theta_{jE} \sim Gamma(a_{\theta_E}, b_{\theta_E})$ and $f_{p_{jg}}(p|g \in G_{jE}) = Beta(0.5, \theta_{jE})$. For some prior probability that $g \in G_{jE}$, η_g, we define the probability $Y_{jg} = y$ as

$$f_{Y_{jg}}(y; N_{jg}, \Theta) = \eta_g \cdot \text{Pr}\left(Y_{jg} = y; N_{jg}, (0.5, \theta_{jE})\right) + (1 - \eta_g)\,\text{Pr}\left(Y_{jg} = y; N_{jg}, (m_{jI}, \theta_{jI})\right)$$
$$\text{for } \Theta = \left\{\eta_g, m_{jI}, \theta_{jI}, \theta_{jE}\right\}. \quad (6)$$

for $\Theta = \{\eta_g, m_{jI}, \theta_{jI}, \theta_{jE}\}$.

To leverage existing biological knowledge of patterns of XCI while simultaneously borrowing information across samples, we model η_g using probit regression. Let X denote a $G \times 4$ design matrix where the columns of X are binary dummy variables for previously-defined consensus XCI states of 'subject', 'mostly subject', 'mostly escape', and 'escape' status. We define $\Phi^{-1}(\boldsymbol{\eta}) = \beta_0 + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\varepsilon_g \sim N(0, \sigma_G^2)$ correspond to independent and identically distributed Gaussian random gene effects with variance σ_G^2 and $\Phi^{-1}(\cdot)$ is the inverse cumulative density function of the standard Gaussian distribution. The posterior probability of escape for gene g and sample j, $PP_{jg} = P(g \in G_{jE}|Y_j)$ can then be derived as:

$$PP_{jg} = \frac{\eta_g \cdot \text{Pr}\left(Y_{jg} = y; N_{jg}, (0.5, \theta_E)\right)}{\eta_g \cdot \text{Pr}\left(Y_{jg} = y; N_{jg}, (0.5, \theta_E)\right) + (1 - \eta_g)\text{Pr}\left(Y_{jg} = y; N_{jg}, (m_{jI}, \theta_{jI})\right)}$$

Priors for m_{jI} are independent flat *Uniform*(0,0.5) distributions. An improper prior is applied for σ_G and a vague independent Gaussian prior $N(0,10)$ for β_0. Priors on $\boldsymbol{\beta}$ preserve the ordinal nature of the consensus state categories, such that $\pi(\beta_1) \sim N(0,10)$, $\pi(\beta_k^*) \sim N^+(0, 10)$, and $\beta_k = \beta_1 + \sum_2^k \beta_k^*$ for $k = 2,3,4$. Hyperpriors on a_{θ_E} and b_{θ_E} are informative priors of $\pi(a_{\theta_E}) \sim Gamma(10,1)$ and $\pi(b_{\theta_E}) \sim Gamma(10,10)$ based on estimates for autosomal genes of eight available normal tissue samples. We also assume $\theta_{jI} \quad \theta_{jE}$ by modeling $\theta_{jI} = \theta_{jE} + \theta_{jE}^*$ and $\pi\left(\theta_{jE}^*\right)$ corresponding to a vague Gamma distribution.

Application to Ovarian Cancer

To investigate patterns of XCI in ovarian cancer, we applied our two-stage XCI analytical methods to all ovarian tumor samples that met our inclusion criteria (Figure 3). Samples were declared significantly skewed at a Bonferroni-adjusted alpha level of 0.05, with the adjustment factor determined by the total number of tested samples. Skewed samples were then carried forward for genic XCI escape analyses. All samples with ≥ 5 genes in $\mathscr{G}_I^*$ and $N_g \geq 10$ were considered to have sufficient data for XCI skewness testing. All observations corresponding to $N_{jg} \geq 10$ were considered to have sufficient gene-level information for genic XCI escape inference, and classification of XCI status of a given gene was determined via a simple decision criterion (i.e., $g \in G_{jE}$ if $PP_{jg} > 0.5$).

Using our multi-sample approach, we again apply MCMC methods to fit the model to the data using two independent chains of 5000 iterations after a burn-in of 5000. To investigate the influence of the various prior definitions on the final results, we ran single-sample analyses using flat and informative priors in addition to our multi-sample approach.

RESULTS

Skewness Test Simulations

As anticipated, statistical power to detect skewed XCI samples was correlated with increased degree of skewness, increased ASE levels, and higher total number of genes tested (Table I). Our test also demonstrated conservative performance with respect to Type I error control, with estimated false positive proportions near 0.01 at the 0.05 α level. This is likely due to distributional differences between the assumed beta-binomial distribution and the true ASE count data, since we applied a parametric bootstrap approach to characterize the null distribution. However, overall statistical power was high for a large proportion of realistic data conditions, with nearly 100% power to detect instances of skewness when $m_I \leq 0.2$ under the various levels of L and N_g under consideration. We also found our method to be robust to the accidental inclusion of true escape genes, particularly when m_I was low (Table S3). We only observed a substantive loss in power for $m_I \leq 0.3$ when 20% of the included genes were simulated to be escaping XCI, with absolute power differences compared to our original simulations (without escape gene contamination) ranging from 0.08 to 0.27.

Escape Gene Analysis Simulations

Performance for escape gene classification was improved as level of skewness and number of reads increased (Table II). Accuracy for classifying escape genes was very high when the informative prior was correctly specified, ranging from 0.72 to 0.94 based on the number of reads and level of sample skewness. Furthermore, the method is robust to prior misspecification, with a reduction in accuracy of only 0.01–0.06 when an informative prior with 10% error was used. The method also performed well with an uninformative prior, especially for highly skewed samples with a large number of reads (accuracy=0.90, sensitivity=0.98, specificity=0.98). Performance was lower for samples with a low level of skewness (m=0.3), driven primarily by lower levels of specificity for escape gene detection.

Data Application – Ovarian Cancer

We applied our two-stage approach to the ASE data estimated from the ovarian cancer dataset. We did not observe any differences in median ASE due to genotyping platform ($p_{610} = p_{2.5M} = 0.44$, $p_{exome} = 0.15$) or RNAseq batch ($p = 0.99$).

CNA filtering

As CNAs are common in ovarian cancer, a large number of tumor samples were excluded due to CNAs overlapping the genes in $\mathscr{G}_I^*$, the set of genes likely to be inactive on the X_i based on prior literature, and were excluded from further analysis. Tumors from 52 patients exhibited overlapping CNAs for all X-linked genes with expression data or lacked ASE data for genes in $\mathscr{G}_I^*$ after CNA segment filtering, and were excluded from further analysis (Figure 3). Many of these excluded samples exhibited aneuploidy, or short or long arm aberrations typical of ovarian cancer[Cheng, et al. 1996; Huang, et al. 2012].

Skewness testing

After filtering based on CNAs, 47 tumor samples were adequate for skewness testing, of which 45 demonstrated significantly skewed XCI after correction for multiple testing (Table S4). Maximum likelihood estimates (from skewness testing) and posterior medians (XCI escape analysis) of m_{jI} were highly concordant across the tumor samples (Spearman $\rho = 0.858$), and indicated trends toward extremely skewed XCI, with $\hat{m}_{jI} < 0.05$ for 34/45 samples under both analyses (Figure 4).

Escape gene analysis

In total, $G = 397$ X chromosome genes were evaluated for XCI escape in at least one tumor sample, of which 318 (80.1%) corresponded to *a priori* information of escape status in **X**. The number of post-filtering ASE gene measurements analyzed per tumor sample ranged from 8 to 142 genes, with a median of 81, while the number of informative samples per gene ranged from 1 to 38 (median = 6). The final data were highly sparse due in combination to low total gene expression and/or extensive CNA overlap in remaining tumor samples, with a missingness rate of approximately 84% across all samples and genes. For all 16 samples with sufficient reads to evaluate XCI escape status for *XIST*, allelic expression ratios were close to zero (Supplemental Figure 1) and estimated posterior probabilities of escape suggested monoallelic expression as expected ($PP_{jg} < 0.05$), providing a proof of principle for our methods. No genes resided in the pseudo-autosomal region after filtering, so could not be evaluated for expected high predicted probabilities of escape.

In comparing posterior escape probabilities across the three analyses, results were overall highly concordant (Table S5). The Spearman correlation coefficients of the PP_{jg} across all samples and genes ranged from 0.61 to 0.86, with the multi-sample results and the flat prior single-sample results being the most correlated. Classifications of XCI escape status based upon a decision threshold of 0.50 were also highly similar, with Cohen's kappa statistics ranging from 0.57 to 0.67. The multi-sample analysis tended to produce more extreme posterior escape probabilities, with approximately 60% of PP_{jg} values < 0.01 or > 0.99. In contrast, the flat and informative prior single-sample analyses, respectively, corresponded to

15.5% and 40% of PP_{jg} values that were similarly extreme. This is likely due to strong prior escape probabilities being inferred by consistent gene-level results across the analyzed samples in the multi-sample analysis. Incidentally, these strong priors would also require greater evidence to identify sporadic instances of discordant gene-level XCI. Results below are reported for the multi-sample analysis using an informative prior.

Concordance with XCI consensus states

Examination of results from the escape gene analysis revealed that genes in $\mathscr{G}_I^*$ demonstrated highly concordant escape analysis results across samples and were estimated to be subject to XCI; similarly, many genes in $\mathscr{G}_E^*$ were estimated to escape from XCI (Figure 5A). Example gene-level XCI escape analysis results comparing allelic expression ratios to PP_{jg} for 38 observations of the gene *BGN*, encoding biglycan, are presented in Figure 5B, which illustrates a high correlation between allelic expression ratio and posterior probability of escape; while *BGN* is thought to undergo XCI, many tumor samples showed evidence of escaping XCI while some tumor tissue samples exhibited mono-allelic expression (Table III). Similarly, Figure 5C shows high correlation between allelic expression ratios and PP_{jg} for the gene *USP9X*, a putative tumor suppressor which is thought to escape XCI, but many tumor samples show evidence of silencing (Table IV). Figure 5D displays the posterior escape probabilities for all genes in $\mathscr{G}_E^*$ and $\mathscr{G}_I^*$ for which at least 10 samples were considered informative for XCI escape analyses (i.e., $N_{jg} \geq 10$), separated by XCI consensus states.

At the sample-level, individual samples also demonstrated high levels of concordance between the most likely genic escape status and the consensus states; on average, 10.3% of *a priori* 'escape/mostly escape' or 'subject/mostly subject' XCI genes were discordant for a given sample (Table S6). Of the discordant genes for a given sample, 58.4% were prior XCI escape genes now classified as subject to XCI.

Discordant genes

To isolate a high-confidence subset of genes with potentially aberrant XCI states across the tumor samples compared to the consensus state, we re-classified gene escape status under greater stringency ($g \in G_{jE}$ if $PP_{jg} > 0.90$, $g \in G_{jI}$ if $PP_{jg} < 0.10$) and identified the subset of genes that corresponded to a gene in $\mathscr{G}_I^*$ or $\mathscr{G}_E^*$ and ≥ 1 tumor sample was predicted to be contrary to that state based on the stringent definition (i.e., $PP_{jg} > 0.90$ and $g \in \mathscr{G}_I^*$). This analysis identified 37 discrepant genes, including 41 observations for 17 unique genes in $\mathscr{G}_I^*$ with evidence of genes in tumor samples putatively escaping XCI (Table III). Review of the allele-specific read counts for these genes revealed allelic expression ratios near 0.5, and 5 genes (*CXorf36, SH3BGRL, ELF4, SLITRK4, TAZ*) have previously been associated with oncogenic properties, and many of these genes reside on the 'q arm' of the X. The expression of many of these genes are positively correlated with *XIST* expression (p<0.05 for all but *DMD, ELF4, FHL1*, and *GABRE*; data not shown). Additionally, 88 tumor observations of 20 unique genes corresponded to genes classified as subject to XCI in the tumor despite belonging to $\mathscr{G}_E^*$ (Table 4). 7 genes (*KAL1, RBBP7, TCEANC, TRAPPC2, DDX3X, USP9X, KDM5C*) have been previously associated with tumor suppression, and

two (*CTPS2*, *RPS4X*) have associations with treatment response or prognosis, and many of these genes reside on the 'p arm' of the X. The expression of all of these were positively associated with *XIST* expression (p<0.05; data not shown). The full gene-level results are presented in Table S7.

For many of the genes not classified into $\mathscr{G}_I^*$ or $\mathscr{G}_E^*$ based on consensus calls across multiple tissue types, our results suggest consistent XCI status in ovarian tumor tissue. For example, many genes previously classified as variable escape across multiple normal tissue types showed strong evidence of undergoing XCI in more than 10 ovarian tumor samples, such as *HCFC1*, *MED14*, *PLS3*, *SHROOM2*. Furthermore, genes previously labeled as 'discordant' across tissues (where the gene was consistently escaping in one tissue but not all tissues) also showed consistent evidence of undergoing XCI in more than 10 ovarian tumor samples, specifically *DIAPH2*, *PIR*, *PNMA3*, *REPS2*, and *TBL1X*. Lastly, among genes that previously had undetermined XCI status, we observed sufficient allele-specific expression in more than 10 ovarian tumor samples, providing evidence to determine XCI status in ovarian tissue. In particular, *FAM127C* and *TMSB4X* show evidence of undergoing XCI, whereas *LOC389906* and *TLR8* show evidence of escape from XCI and variable escape across ovarian tumor tissues, respectively.

DISCUSSION

We have developed a framework for the multi-omic analysis of XCI, have evaluated the performance of our method in simulation studies, and have applied our method to estimate genes that escape XCI in ovarian cancer. Integrating transcriptome sequence data, genome-wide genotyping, and copy-number segmentation results, we also characterized transcriptional XCI patterns in ovarian carcinoma samples with X-linked allelic expression data. In general, we observed allelic expression patterns consistent with the consensus gene-level XCI classifications, particularly genes subject to XCI. Across tumors, most samples corresponded to predicted active Xi transcription for only a small proportion of genes that were *a priori* anticipated to be inactive, indicative of sporadic XCI reactivation for discordant genes. Similarly, only a small proportion of genes that were a priori anticipated to escape XCI were predicted to be inactivated for most tumor samples, suggesting either deactivation of Xi, or possibly tissue-specific differences in XCI patterns.

The two-stage methods developed here to assess XCI skewness and gene-level XCI escape status for individual tumor samples have many advantages compared to previous approaches to estimate XCI escape genes. Previous approaches have been statistically simplistic, and have not utilized formal statistical tests or sophisticated probability models, and instead have implemented thresholds of allelic imbalance for escape status classification. We provide a statistical framework for modeling allele-specific expression data in the context of X chromosome inactivation based on known biological properties. Because most genes undergo silencing yet some escape inactivation, we assume a mixture distribution to account for the expected distributions under both of these states. Due to the over-dispersion typically present in expression data, we assume a beta-binomial model for this mixture distribution. Our framework is flexible, and could easily be extended as more is known about X chromosome biology; for example, 'variable escape' mixture state could be include, or the

choice of reference XCI status could be updated based on domain or tissue-specific knowledge. Furthermore, this method could be applied to any study of XCI with available RNAseq and genotype data, and is not limited to studies of ovarian cancer. For studies of phenotypes affecting both males and females, the method could be extended to include male samples, and future directions involve incorporating total expression data to improve inference.

Using this biologically-driven mixture distribution to model the XCI process and account for two possible XCI states, we first developed a formal statistical test for skewness (based on a likelihood ratio test), allowing us to determine not only which samples exhibit skewed XCI, but also quantify the level of that skewness for a given sample. Furthermore, our second stage escape gene analysis provides estimates of the posterior probability of escape for each gene, for each individual sample; such estimates have not been provided with other approaches, and can aid in follow-up studies to connect the degree of escape from XCI with clinical characteristics. Because the process of XCI occurs in multiple latent states (e.g., escape and silenced) where prior knowledge about these states is available for many genes and a posterior probability of state membership is desired, the Bayesian paradigm is a natural fit for XCI escape gene analysis, which has not previously been utilized.

Additionally, our integrative genomic approach based on multiple genomic data types uses a direct measure of XCI (allele-specific expression data), as opposed to indirect measures such as total expression data or methylation data, and can estimate the degree of XCI dysregulation independent from copy number alterations. Furthermore, the use of ASE data mitigates the presence of batch effects, as the proportion of reads mapping to a particular allele is not dependent on the total coverage. Most importantly, our approach does not rely on analogous male samples for comparison, which are not available for female-specific disorders (such as gynecological cancers)—the phenotypes for which study of XCI is often most relevant. While in this study we applied these methods to a study of ovarian cancer, these methods are not cancer-specific and would be applicable to any study of XCI in females.

We evaluated the performance of our proposed two-stage approach in a series of simulations studies. We examined the type I error and power of the stage 1 test for skewness, and observed that the type I error of the likelihood ratio test is somewhat conservative, likely due to the use of a bootstrap approximation. However, the conservative nature of the test is offset by the extremely high power of the approach to detect moderate to high levels of skewness (consistent with what was observed in the ovarian tumor data). We then evaluated the classification performance of our approach to detect genes that escape from XCI. Accuracy, sensitivity, and specificity were high, especially under the use of an informative prior derived from the consensus states published in prior literature, although the method was also robust to the choice of a flat prior or misspecification of an informative prior, with performance similar for highly skewed samples with even low coverage.

In the stage 1 test for skewness, we observed a large degree of skewed XCI across tumors. These results are consistent with the notion that the clonal nature of tumors would correspond to relative homogeneity of the Xi in the corresponding tissue sample, and

underscore the viability of exploring XCI patterns in tumor expression data. Although not evaluated here, a high degree of skewness may also be observed in normal ovarian tissue, because silencing of the inactive X occurs early in development at the time of gastrulation[van den Berg, et al. 2009], prior to the differentiation of Müllerian tissues. Furthermore, the high degree of tumor skewness is also consistent with observed increased levels of skewed XCI in lymphocytes of ovarian cancer patients compared to controls[Buller, Sood, Lallas, Buekers, & Skilling 1999; Lose, Duffy, Kay, Kedda, & Spurdle 2008], although subsequent studies have not observed this phenomenon in blood[Manoukian, et al. 2013]. This also suggests a lack of patients in our study with substantial chromosome-wide reactivation of the inactive X.

In the stage 2 analysis of gene-level XCI escape status, we identified 17 genes escaping XCI contrary to consensus patterns of normal XCI, indicating potential somatic alterations to XCI status, many of which have noted associations with oncogenic properties in the literature. Expression of most of these genes is positively correlated with *XIST* expression. Oncogenes showing evidence of unexpected escape from XCI in ovarian tumors highlights their potential for involvement in ovarian tumor initiation and progression due to increased gene dosage. Four tumors indicated potential aberrant XCI escape of *CXorf36*, which is highly expressed in ovarian tissue and encodes a protein of unknown function. A recent study of renal carcinoma indicated upregulated expression and 5′ regional hypomethylation of *CXorf36,* while similar results were observed in a study of polycystic ovarian syndrome[Wang, et al. 2014]. *SLITRK4* is a member of the SLITRK transmembrane protein family that modulate neurite outgrowth, and overexpression of *SLITRK4* mRNA and protein products have been noted in primary uterine leiomyosarcomas[Davidson, et al. 2014]. *ELF4*, also known as *MEF*, is an ETS-family transcriptional activator that has been associated with both tumor suppression and oncogenic function, and *ELF4* overexpression has been shown to increase the oncogenic properties of ovarian cancer cells[Yao, et al. 2007]. Similarly, *SH3BGRL* demonstrates oncogenic activity in mice but tumor suppressive behavior in humans[Wang, et al. 2015], with upregulation of SH3BGRL protein observed in triple-negative breast cancer, a subtype of breast cancer with many biological, epidemiological and etiological similarities to high-grade serous ovarian cancer [Muniz Lino, et al. 2014]. Furthermore, tafazzin, encoded by *TAZ*, has been shown to promote the epithelial-mesenchymal transition and progression and tumorigenesis of many cancer types, including ovarian cancer[Chen, Xie, Huang, & Yang 2016; Moroishi, Hansen, & Guan 2015].

In contrast to the above listed genes, some tumor samples indicated XCI silencing of 20 escape genes, including many potential tumor suppressors. This high level of tumor suppressor genes that exhibit unanticipated silencing in ovarian tumors is noteworthy because of their potential for involvement in ovarian tumor initiation and progression due to reduced gene dosage. For instance, 6 tumors exhibited aberrant deactivation of *KAL1*, Kallman Syndrome sequence 1 gene, has been attributed tumor suppressive properties in studies of squamous cell carcinoma[Liu, Cao, Chen, Xu, & Zhang 2015] and hepatocellular carcinoma[Tanaka, et al. 2015]. *RBBP7* protein RBBP7 (retinoblastoma binding protein 7) has been shown to have tumor suppressive properties by interacting with NKX6.1 in repressing vimentin and N-cadherin expression[Li, et al. 2016]. Notably, aberrant silencing of X-linked genes has previously been observed in studies of XCI in breast tumor cell

lines[Chaligne, et al. 2015], indicating epigenetic modification of XCI is possible in either direction and may act as a mechanism of disrupting tumor suppressor activity. *TCEANC* is thought to interact with *BRCA1* in ovarian tumor suppression[Hill, et al. 2014], and this gene was aberrantly silenced in seven samples. *TRAPPC2* protein, trafficking protein complex 2, also known as *MIP-2A*, may inhibit tumor proliferation and could be a potential drug target[Tokunaga, et al. 2013]. *DDX3X*, encoding DEAD-box helicase 3 X-linked, is a putative tumor suppressor, and has also been associated with escape from XCI that could contribute to male sex bias for some cancers[Bol, Xie, & Raman 2015; Dunford, et al. 2016]. *USP9X*, encoding ubiquitin specific peptidase 9 X-linked, has been linked to tumor suppression in pancreatic and colorectal cancer[Jo, Kim, Yoo, & Lee 2016; Perez-Mancera, et al. 2012]. Interestingly, all eight samples evaluated for *KDM5C* showed evidence of inactivation. *KDM5C* encodes a histone lysine demethylase (KDM), and studies have suggested KDMs may act as both oncogenes and tumor suppressors for a variety of cancers[Cloos, Christensen, Agger, & Helin 2008]. *KDM5C* contributes to ovarian development[Sun, et al. 2016], is a candidate tumor suppressor in renal cancer[Dalgliesh, et al. 2010], and escape from XCI of *KDM5C* is also associated with cancer male bias[Dunford, et al. 2016].

Furthermore, tumor-specific profiles of XCI may not only provide additional information in identifying novel cancer genes in ovarian cancer, but could prove clinically useful in personalized medicine. Analyses of *XIST* expression in ovarian carcinoma tissues demonstrated significant correlations with chemotherapeutic response[Huang, et al. 2002]. In our study, two prior escape genes that showed evidence of silencing in the ovarian tumor (*CTPS2* and *RPS4X*) have previously been shown to be associated with treatment response and prognosis. In cell lines, increased expression of *CTPS2* was associated with reduced platinum sensitivity (carboplatin and cisplatin)[Gamazon, et al. 2011], and low expression of *RPS4X* was associated with poor prognosis in serous ovarian cancer [Tsofack, et al. 2013]. Furthermore, *TRAPPC2* and *DDX3X* have also been identified as possible drug targets[Bol, Xie, & Raman 2015; Tokunaga, et al. 2013].

Interestingly, we also observed regional patterns regarding discrepant XCI states outside of the pseudo-autosomal regions. Many of the genes that indicated potential aberrant re-activation compared to the consensus state (including those with oncogenic properties) are located on the short (p) arm of the X chromosome, whereas many of the genes showing evidence of aberrant silencing compared to the consensus state (including many with tumor suppressive properties) are located on the long (q) arm of the X chromosome. This suggests that re-activation may be more likely on the p arm and de-activation may be more common on the q-arm, although further studies are needed to validate this finding and determine its significance.

There may also be relevant genes of interest to ovarian cancer that do not correspond to active or inactive XCI consensus states. *TBL1X*, denoted "discordant" in Balaton et al., was recently identified by Chaligne et al.[Chaligne, et al. 2015] to aberrantly reactivate on the Xi in breast cancer cell line ZR-75-1, resulting in increased gene and protein expression dosage. In our analyses, 23 tumors had available *TBL1X* ASE data for XCI escape analysis, of which two corresponded to posterior escape probabilities > 0.99. These two observations of *TBL1X*

escaping XCI may be attributable to either somatic alteration in otherwise silenced Xi alleles, or normal inter-sample variability of XCI for *TBL1X* in ovarian epithelial cells. These findings highlight the importance of tissue-specific expectations of gene-level XCI status as well as paired tumor-normal tissue analyses in cancer studies of XCI, which may better inform likely instances of aberrant XCI activity. These analyses relied on consensus XCI states across multiple tissues for tumor comparisons, but due to potential tissue-specificity of ovarian XCI patterns, future studies to examine XCI states in normal ovarian tissue are warranted.

In addition to lack of paired normal expression data to characterize normal ovarian XCI patterns, there are other limitations to this study. For instance, gene-level inference was limited to only a fraction of the total expressed X-linked genes due to the additional data burden of RNA-Seq reads overlapping eSNPs, and many genes corresponded to zero read counts for ASE. Additional integration of available total expression may make the inclusion of more genes in XCI analysis possible; however, analytical complications arise with the correlation structure of gene co-expression as well as substantial transcriptomic dysregulation in the tumor environment. Furthermore, the detection of skewed XCI samples is dependent upon the assumption that the XCI status of the majority of presumably inactive genes is preserved, as the skewness testing we conducted cannot discriminate between complete reactivation of the Xi and high cellular heterogeneity of XCI. Consequently, tumor samples that failed skewness testing may alternatively correspond to full epigenetic erosion of the Barr body. Moreover, allelic imbalance for a particular gene could be driven by factors unrelated to XCI, such as *cis*-acting regulatory genetic variation, and genes indicating potential aberrant deactivation on the Xi in tumor samples would require confirmation with paired normal tissue to exclude this possibility. Further integration of epigenetic markers, such as promoter DNA methylation and histone modifications, may provide supporting evidence of genic XCI status and strengthen multi-omic XCI analyses. Finally, our analysis strategies were dependent upon assumed normal copy number states for the assessed genes, and analytical methods that could incorporate the presence of CNAs and LOH would be of both biological and clinical interest. Future evaluation of XCI in cancer would especially benefit from single-cell analyses, which would be able to circumvent issues of uncertainty attributable to potential cellular homogeneity of Xi as well as normal tissue contamination.

CONCLUSIONS

We demonstrate a novel framework that shows dysregulation of XCI in individual ovarian tumors for a number of genes across the X chromosome. While general patterns of XCI in ovarian cancers were largely consistent with previously identified XCI consensus states, we identified discrepant genic XCI status classifications for a number of X-linked genes, including many with potential biological relevance to tumorigenesis and tumor suppression. These results provide initial evidence of potential somatic alteration to XCI in ovarian cancer and highlight an additional mechanism of gene expression dysregulation. Future studies using paired tumor-normal tissue analysis, multi-omic integration of epigenetic data, and single-cell sequencing may further elucidate the degree of susceptibility of XCI to dysregulation as well as its role in the risk, progression, and treatment resistance of ovarian cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Balaton BP, Cotton AM, Brown CJ. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. Biology of sex differences. 2015; 6:35. [PubMed: 26719789]

Barr ML, Moore KL. Chromosomes, sex chromatin, and cancer. Proceedings Canadian Cancer Conference. 1957; 2:3–16. [PubMed: 13437159]

Berletch JB, Yang F, Xu J, Carrel L, Disteche CM. Genes that escape from X inactivation. Hum Genet. 2011; 130(2):237–45. [PubMed: 21614513]

Bol GM, Xie M, Raman V. DDX3, a potential target for cancer treatment. Mol Cancer. 2015; 14:188. [PubMed: 26541825]

Buller RE, Sood AK, Lallas T, Buekers T, Skilling JS. Association between nonrandom X-chromosome inactivation and BRCA1 mutation in germline DNA of patients with ovarian cancer. J Natl Cancer Inst. 1999; 91(4):339–46. [PubMed: 10050867]

Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker MA, Guo J, Li P, Riddell A. Stan: a probabilistic programming language. Journal of Statistical Software. 2015

Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature. 2005; 434(7031):400–4. [PubMed: 15772666]

Chaligne R, Heard E. X-chromosome inactivation in development and cancer. FEBS letters. 2014; 588(15):2514–22. [PubMed: 24937141]

Chaligne R, Popova T, Mendoza-Parra MA, Saleem MA, Gentien D, Ban K, Piolot T, Leroy O, Mariani O, Gronemeyer H, et al. The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. Genome research. 2015; 25(4):488–503. [PubMed: 25653311]

Chen G, Xie J, Huang P, Yang Z. Overexpression of TAZ promotes cell proliferation, migration and epithelial-mesenchymal transition in ovarian cancer. Oncol Lett. 2016; 12(3):1821–1825. [PubMed: 27588129]

Cheng PC, Gosewehr JA, Kim TM, Velicescu M, Wan M, Zheng J, Felix JC, Cofer KF, Luo P, Biela BH, et al. Potential role of the inactivated X chromosome in ovarian epithelial tumor development. J Natl Cancer Inst. 1996; 88(8):510–8. [PubMed: 8606379]

Cloos PA, Christensen J, Agger K, Helin K. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. Genes & development. 2008; 22(9):1115–40. [PubMed: 18451103]

Cotton AM, Ge B, Light N, Adoue V, Pastinen T, Brown CJ. Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. Genome biology. 2013; 14(11):R122. [PubMed: 24176135]

Cotton AM, Price EM, Jones MJ, Balaton BP, Kobor MS, Brown CJ. Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Human molecular genetics. 2015; 24(6):1528–39. [PubMed: 25381334]

Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010; 463(7279):360–3. [PubMed: 20054297]

Davidson B, Abeler VM, Forsund M, Holth A, Yang Y, Kobayashi Y, Chen L, Kristensen GB, Shih Ie M, Wang TL. Gene expression signatures of primary and metastatic uterine leiomyosarcoma. Human pathology. 2014; 45(4):691–700. [PubMed: 24485798]

Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nature methods. 2013; 10(1):5–6. [PubMed: 23269371]

Dunford A, Weinstock DM, Savova V, Schumacher SE, Cleary JP, Yoda A, Sullivan TJ, Hess JM, Gimelbrant AA, Beroukhim R, et al. Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. Nat Genet. 2016

Gamazon ER, Im HK, O'Donnell PH, Ziliak D, Stark AL, Cox NJ, Dolan ME, Huang RS. Comprehensive evaluation of the contribution of X chromosome genes to platinum sensitivity. Molecular cancer therapeutics. 2011; 10(3):472–80. [PubMed: 21252287]

Henderson BE, Feigelson HS. Hormonal carcinogenesis. Carcinogenesis. 2000; 21(3):427–33. [PubMed: 10688862]

Hill SJ, Rolland T, Adelmant G, Xia X, Owen MS, Dricot A, Zack TI, Sahni N, Jacob Y, Hao T, et al. Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. Genes & development. 2014; 28(17):1957–75. [PubMed: 25184681]

Huang KC, Rao PH, Lau CC, Heard E, Ng SK, Brown C, Mok SC, Berkowitz RS, Ng SW. Relationship of XIST expression and responses of ovarian cancer to chemotherapy. Molecular cancer therapeutics. 2002; 1(10):769–76. [PubMed: 12492109]

Huang RYJ, Chen GB, Matsumura N, Lai HC, Mori S, Li J, Wong MK, Konishi I, Thiery JP, Goh L. Histotype-specific copy-number alterations in ovarian cancer. BMC Med Genomics. 2012; 5:47. [PubMed: 23078675]

Jo YS, Kim MS, Yoo NJ, Lee SH. USP9X, a Putative Tumor Suppressor Gene, Exhibits Frameshift Mutations in Colorectal Cancers. Pathol Oncol Res. 2016

Li HJ, Yu PN, Huang KY, Su HY, Hsiao TH, Chang CP, Yu MH, Lin YW. NKX6.1 functions as a metastatic suppressor through epigenetic regulation of the epithelial-mesenchymal transition. Oncogene. 2016; 35(17):2266–78. [PubMed: 26257059]

Liu J, Cao W, Chen W, Xu L, Zhang C. Decreased expression of Kallmann syndrome 1 sequence gene (KAL1) contributes to oral squamous cell carcinoma progression and significantly correlates with poorly differentiated grade. J Oral Pathol Med. 2015; 44(2):109–14. [PubMed: 25060050]

Lose F, Duffy DL, Kay GF, Kedda MA, Spurdle AB. Skewed X chromosome inactivation and breast and ovarian cancer status: evidence for X–linked modifiers of BRCA1. J Natl Cancer Inst. 2008; 100(21):1519–29. [PubMed: 18957670]

Manoukian S, Verderio P, Tabano S, Colapietro P, Pizzamiglio S, Grati FR, Calvello M, Peissel B, Burn J, Pensotti V, et al. X chromosome inactivation pattern in BRCA gene mutation carriers. Eur J Cancer. 2013; 49(5):1136–41. [PubMed: 23146957]

Mclachlan GJ. On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture. Applied Statistics-Journal of the Royal Statistical Society Series C. 1987; 36(3):318–324.

Moroishi T, Hansen CG, Guan KL. The emerging roles of YAP and TAZ in cancer. Nat Rev Cancer. 2015; 15(2):73–9. [PubMed: 25592648]

Muniz Lino MA, Palacios-Rodriguez Y, Rodriguez-Cuevas S, Bautista-Pina V, Marchat LA, Ruiz-Garcia E, Astudillo-de la Vega H, Gonzalez-Santiago AE, Flores-Perez A, Diaz-Chavez J, et al. Comparative proteomic profiling of triple-negative breast cancer reveals that up-regulation of RhoGDI-2 is associated to the inhibition of caspase 3 and caspase 9. Journal of proteomics. 2014; 111:198–211. [PubMed: 24768906]

Pageau GJ, Hall LL, Ganesan S, Livingston DM, Lawrence JB. The disappearing Barr body in breast and ovarian cancers. Nat Rev Cancer. 2007; 7(8):628–33. [PubMed: 17611545]

Perez-Mancera PA, Rust AG, van der Weyden L, Kristiansen G, Li A, Sarver AL, Silverstein KA, Grutzmann R, Aust D, Rummele P, et al. The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. Nature. 2012; 486(7402):266–70. [PubMed: 22699621]

Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010; 464(7289):768–72. [PubMed: 20220758]

Porzio GC, Ragozini G. On the stochastic ordering of folded binomials. Statistics & Probability Letters. 2009; 79(9):1299–1304.

Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, et al. The DNA sequence of the human X chromosome. Nature. 2005; 434(7031):325–37. [PubMed: 15772651]

Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010; 31(1):27–36. [PubMed: 19752007]

Sharp AJ, Stathaki E, Migliavacca E, Brahmachary M, Montgomery SB, Dupre Y, Antonarakis SE. DNA methylation profiles of human active and inactive X chromosomes. Genome Res. 2011; 21(10):1592–600. [PubMed: 21862626]

Sun W. A statistical framework for eQTL mapping using RNA-seq data. Biometrics. 2012; 68(1):1–11. [PubMed: 21838806]

Sun YX, Zhang YX, Zhang D, Xu CM, Chen SC, Zhang JY, Ruan YC, Chen F, Zhang RJ, Qian YQ, et al. XCI-escaping gene KDM5C contributes to ovarian development via downregulating miR-320a. Human genetics. 2016

Tanaka Y, Kanda M, Sugimoto H, Shimizu D, Sueoka S, Takami H, Ezaka K, Hashimoto R, Okamura Y, Iwata N, et al. Translational implication of Kallmann syndrome-1 gene expression in hepatocellular carcinoma. Int J Oncol. 2015; 46(6):2546–54. [PubMed: 25892360]

Tokunaga M, Shiheido H, Tabata N, Sakuma-Yonemura Y, Takashima H, Horisawa K, Doi N, Yanagawa H. MIP-2A is a novel target of an anilinoquinazoline derivative for inhibition of tumour cell proliferation. PLoS One. 2013; 8(9):e76774. [PubMed: 24098805]

Tsofack SP, Meunier L, Sanchez L, Madore J, Provencher D, Mes-Masson AM, Lebel M. Low expression of the X-linked ribosomal protein S4 in human serous epithelial ovarian cancer is associated with a poor prognosis. BMC Cancer. 2013; 13:303. [PubMed: 23800275]

Tukiainen T, Villani A-C, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, et al. Landscape of X chromosome inactivation across human tissues. bioRxiv. 2016

van den Berg IM, Laven JSE, Stevens M, Jonkers I, Galjaard R-J, Gribnau J, Hikke van Doorninck J. X Chromosome Inactivation Is Initiated in Human Preimplantation Embryos. American Journal of Human Genetics. 2009; 84(6):771–779. [PubMed: 19481196]

Wang H, Liu B, Al-Aidaroos AQ, Shi H, Li L, Guo K, Li J, Tan BC, Loo JM, Tang JP, et al. Dual-faced SH3BGRL: oncogenic in mice, tumor suppressive in humans. Oncogene. 2015

Wang XX, Wei JZ, Jiao J, Jiang SY, Yu DH, Li D. Genome-wide DNA methylation and gene expression patterns provide insight into polycystic ovary syndrome development. Oncotarget. 2014; 5(16):6603–10. [PubMed: 25051372]

Winham SJ, Pirie A, Chen YA, Larson MC, Fogarty ZC, Earp MA, Anton-Culver H, Bandera EV, Cramer D, Doherty JA, et al. Investigation of Exomic Variants Associated with Overall Survival in Ovarian Cancer. Cancer Epidemiol Biomarkers Prev. 2016; 25(3):446–54. [PubMed: 26747452]

Yao JJ, Liu Y, Lacorazza HD, Soslow RA, Scandura JM, Nimer SD, Hedvat CV. Tumor promoting properties of the ETS protein MEF in ovarian cancer. Oncogene. 2007; 26(27):4032–7. [PubMed: 17213815]

Yau C, Mouradov D, Jorissen RN, Colella S, Mirza G, Steers G, Harris A, Ragoussis J, Sieber O, Holmes CC. A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data. Genome biology. 2010; 11(9):R92. [PubMed: 20858232]
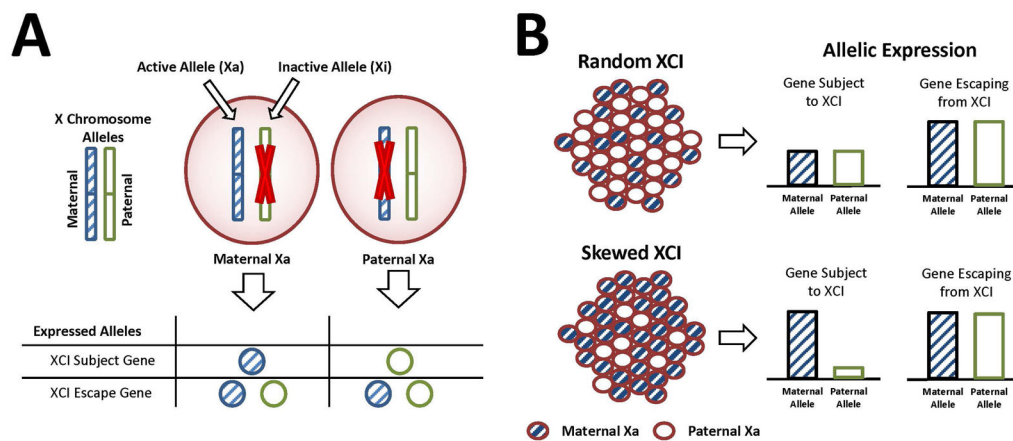
Figure 1.
Measuring XCI with allele expression. **(A)** For a given cell, either the maternal or paternal X is inactivated (top); for XCI Subject genes, only the active copy (Xa) is expressed, whereas for XCI Escape Genes, both the Xa and Xi are expressed (bottom). **(B)** In collections of cells under random XCI, the maternal Xa and paternal Xa are in equal proportions, so genes subject to XCI show no allelic expression imbalance compared to genes escaping XCI (top). Under skewed XCI, either the maternal Xa or paternal Xa is present in higher proportions, leading to allelic expression imbalance for genes subject to XCI (bottom).
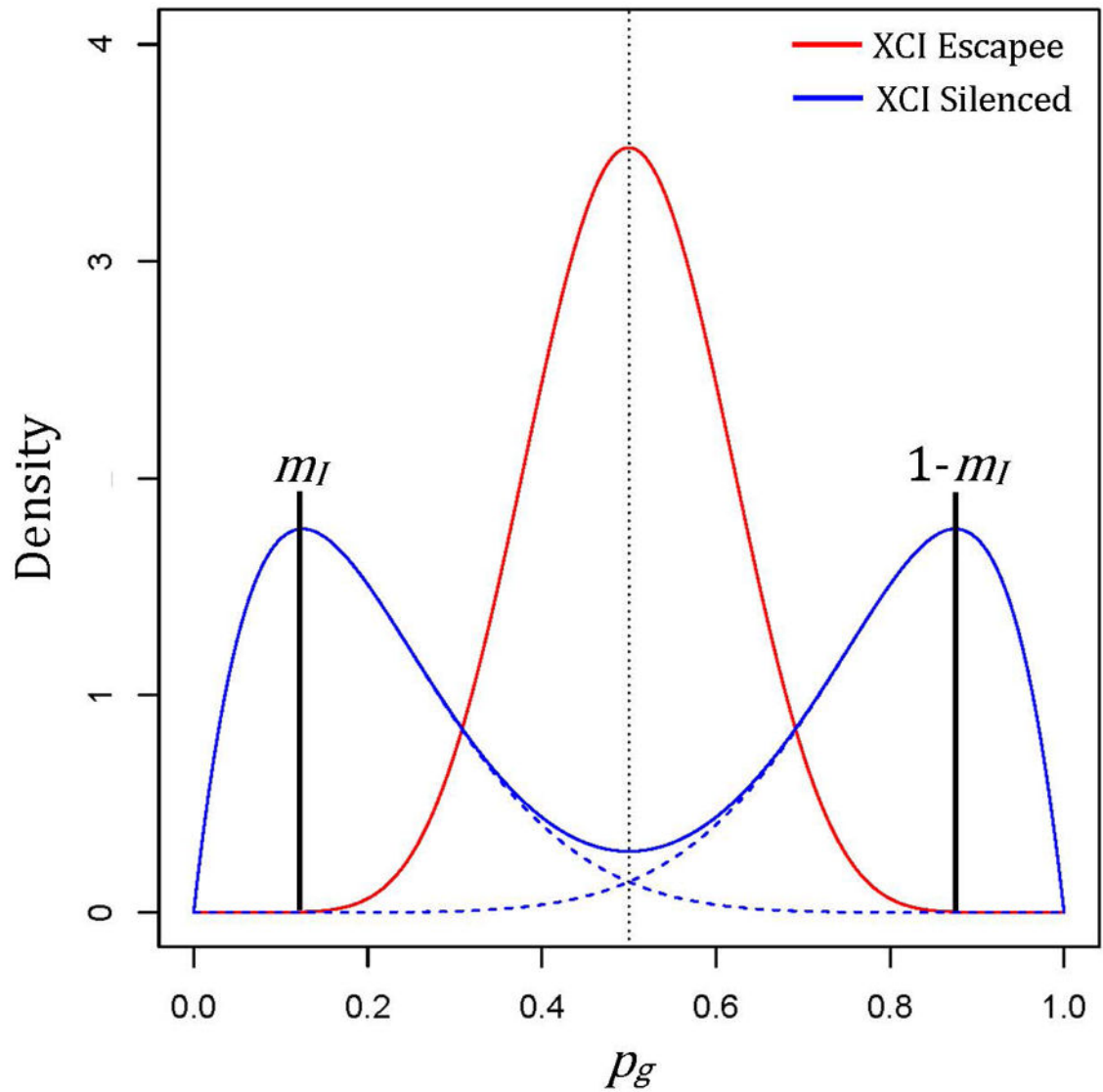
Figure 2.
Depiction of beta-mixtures corresponding to underlying allelic expression ratios (AERs) for genes escaping XCI (red) or silenced by XCI (blue), with degree of skewness (m_I) indicated by vertical black lines at the respective modes of the component symmetric beta distributions (dashed blue lines).
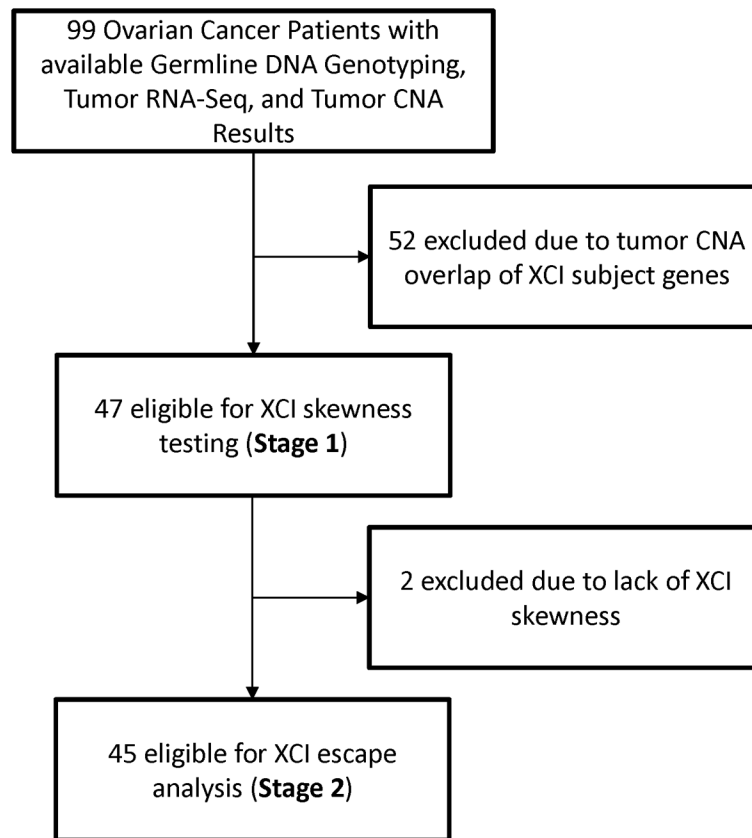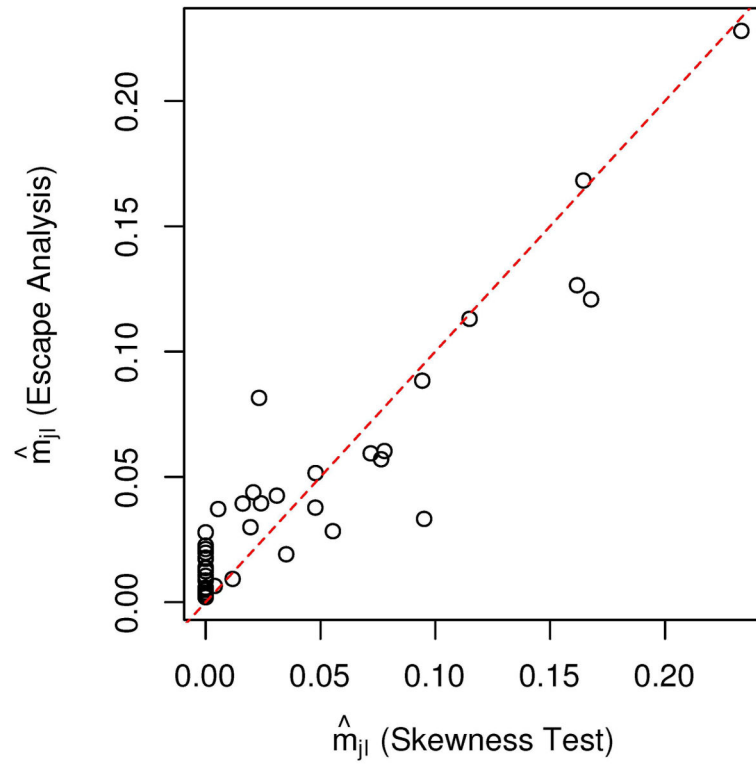
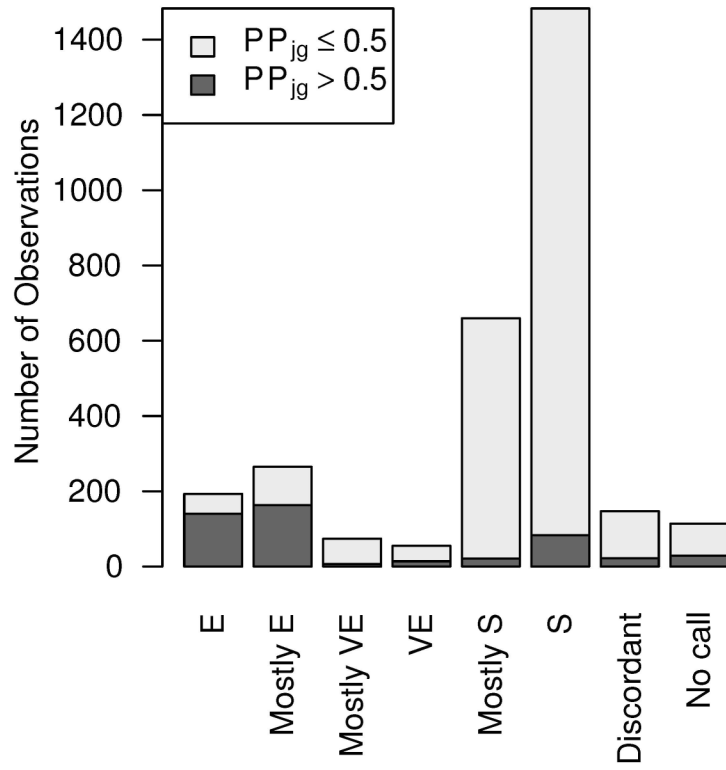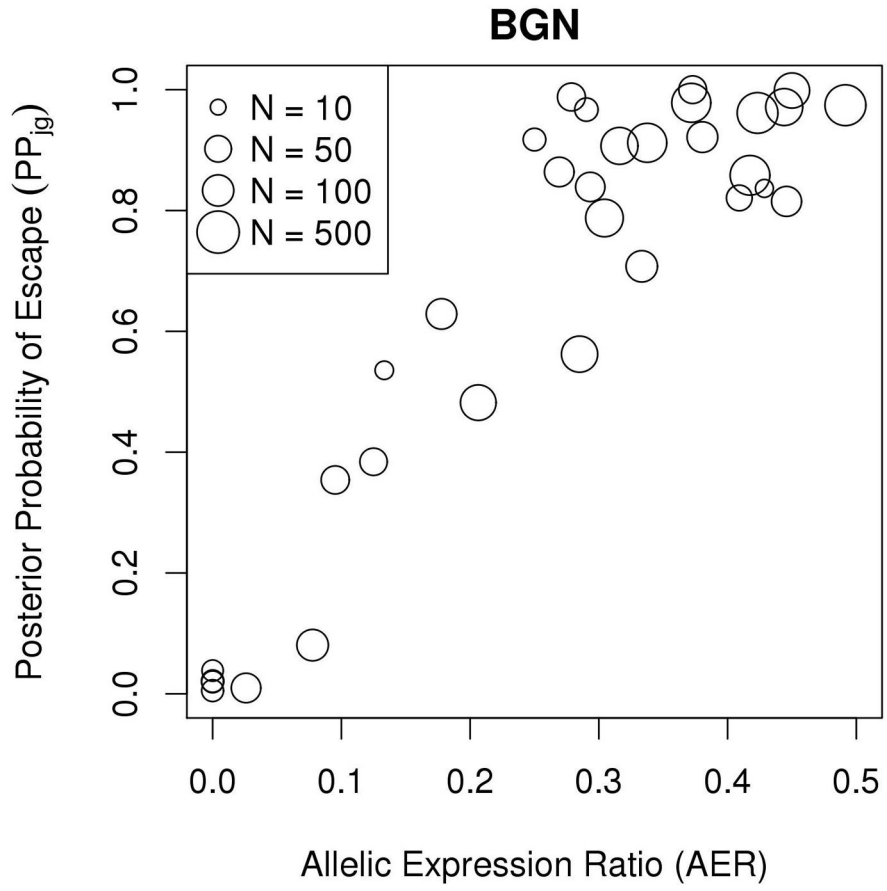**Figure 3.**
Study Flowchart.

Figure 4.
Scatterplot of point estimates of m_{jl} from the skewness test (x-axis) compared with posterior medians from the XCI escape analyses (y-axis) for all 45 skewed XCI ovarian samples.
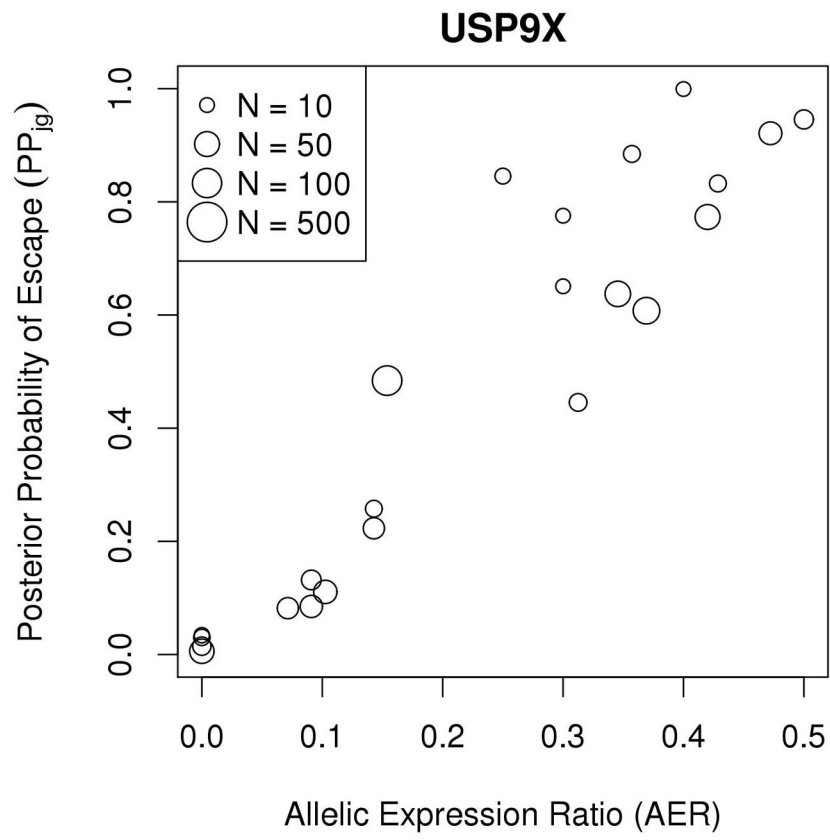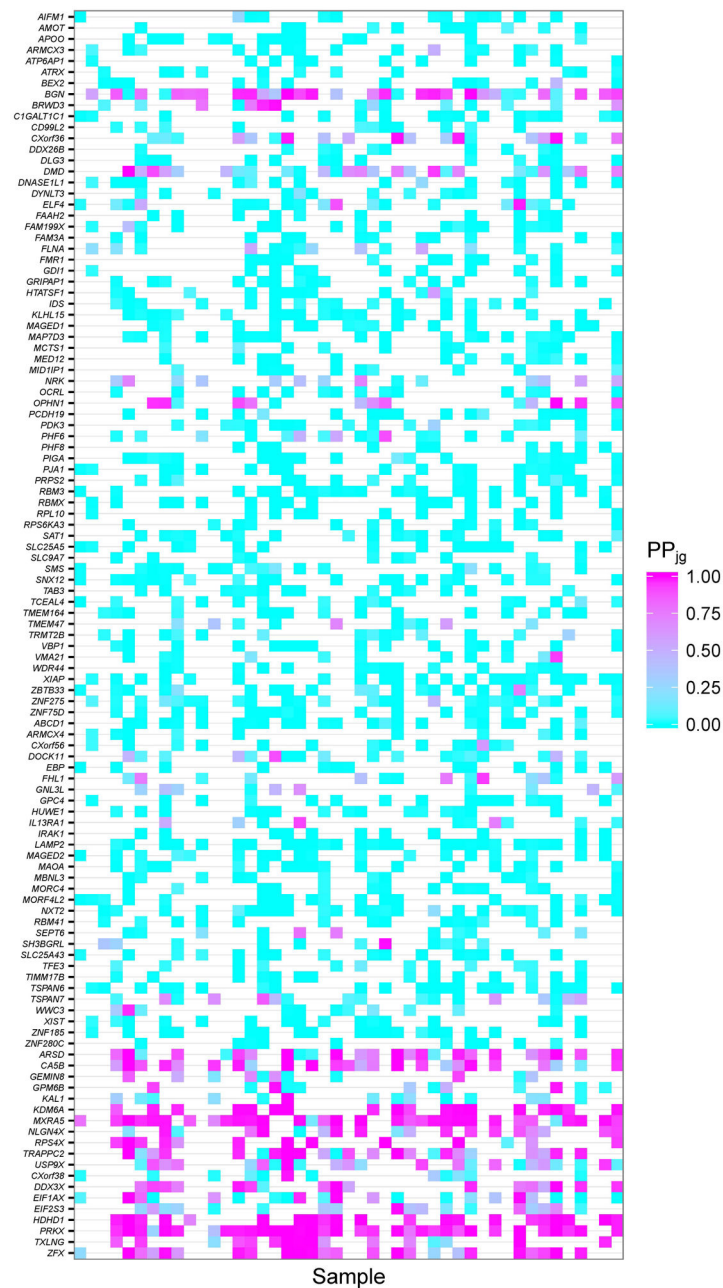
BGN

USP9X

Figure 5.
(**A**) Stacked barplots of XCI escape analysis results for tumor tissue samples, separated by XCI consensus status. 'E'=escape, 'VE'=variable escape, and 'S'=subject, as in Balaton et al. (**B**) Scatterplot of AERs (x-axis) compared to XCI posterior escape probabilities (PP_{jg}) for XCI subject gene *BGN*. Sizes of plotted points are indicative of total number of allele-specific reads (N_{jg}). (**C**) Scatterplot of AERs (x-axis) compared to XCI posterior escape probabilities (PP_{jg}) for XCI escape gene *USP9X*. Sizes of plotted points are indicative of total number of allele-specific reads (N_{jg}). (**D**) Tileplot of the posterior escape probabilities organized by XCI consensus state (Escape, Mostly Escape, Mostly Subject, Subject), with each column corresponding to a unique sample. Coloration is correspondent to calculated

PP$_{jg}$, with gradient of cyan-to-magenta corresponding to the spectrum of low-to-high values of *PP$_{jg}$* on the scale of 0.00 – 1.00. For samples where there were no data for a given gene, the tile is left blank.

Table I

Type I error and power of the XCI skewness test under various data conditions.

| N_g | L | Type I Error | Power | | | | |
		$m = 0.50$	$m = 0.40$	$m = 0.30$	$m = 0.20$	$m = 0.10$
10	10	0.013	0.073	0.484	0.955	1.000
	20	0.006	0.084	0.718	1.000	1.000
	30	0.006	0.101	0.848	1.000	1.000
20	10	0.010	0.079	0.678	0.999	1.000
	20	0.016	0.107	0.901	1.000	1.000
	30	0.007	0.117	0.966	1.000	1.000
30	10	0.016	0.109	0.777	0.999	1.000
	20	0.019	0.128	0.963	1.000	1.000
	30	0.017	0.139	0.989	1.000	1.000

Table II

Simulation results for XCI escape gene analysis. Mean sensitivity, specificity, and accuracy are calculated across the 100 replications and defined with respect to identifying escape genes based upon $PP_{Ig} > 0$. Performance measures for discordant genes under the informative prior are based upon the subset of 10 genes with inaccurate prior definitions.

N_g	m_I	Flat Prior			Informative Prior			Informative Prior with Errors					
								Overall			Discordant Genes		
		Acc	Sens	Spec	Acc	Sens	Spec	Acc	Sens	Spec	Acc	Sens	Spec
10	0.30	0.48	0.82	0.48	0.72	0.69	0.69	0.67	0.64	0.68	0.20	0.20	0.19
	0.20	0.72	0.80	0.72	0.81	0.80	0.80	0.77	0.77	0.77	0.44	0.56	0.31
	0.10	0.75	0.93	0.75	0.86	0.84	0.84	0.83	0.90	0.82	0.66	0.79	0.52
20	0.30	0.64	0.75	0.64	0.77	0.87	0.75	0.73	0.68	0.74	0.28	0.36	0.21
	0.20	0.76	0.88	0.76	0.86	0.94	0.85	0.82	0.84	0.82	0.48	0.69	0.26
	0.10	0.86	0.97	0.86	0.92	0.98	0.90	0.90	0.96	0.89	0.76	0.93	0.59
30	0.30	0.68	0.77	0.76	0.78	0.88	0.76	0.75	0.69	0.75	0.26	0.35	0.16
	0.20	0.80	0.92	0.92	0.87	0.96	0.86	0.84	0.89	0.83	0.50	0.75	0.25
	0.10	0.90	0.98	0.98	0.94	0.98	0.93	0.92	0.97	0.91	0.83	0.95	0.72

Table III

Summary of XCI analysis results for genes thought to undergo XCI with strong evidence of escape (PP>0.90) from XCI in at least one ovarian tumor sample.

Gene	Cytoband (Ensembl)[a]	Total number of tumors evaluated	Proportion of tumors escaping XCI (N escaping/N total)[b]	Consensus State[Balaton, Cotton, & Brown 2015]	Known Oncogene or Tumor Suppressor (reference)
CLCN4	Xp22.2	9	13% (1/8)	Mostly S[c]	
WWC3	Xp22.2	11	14% (1/7)	Mostly S	
CXorf23	Xp22.12	6	50% (2/4)	S[d]	
DMD	Xp21.1	27	33% (3/9)	S	
CXorf36	Xp11.3	19	50% (4/8)	S	Oncogene[Wang, et al. 2014]
OPHN1	Xq12	14	63% (5/8)	S	
BRWD3	Xq21.1	15	22% (2/9)	S	
SH3BGRL	Xq21.1	11	11% (1/9)	Mostly S	Oncogene/tumor suppressor[Muniz Lino, et al. 2014; Wang, et al. 2015]
DOCK11	Xq24	18	7% (1/14)	Mostly S	
IL13RA1	Xq24	10	14% (1/7)	Mostly S	
ELF4	Xq26.1	19	8% (1/13)	S	Oncogene/tumor suppressor[Yao, et al. 2007]
FHL1	Xq26.3	14	17% (1/6)	Mostly S	
SLITRK4	Xq27.3	6	25% (1/4)	S	Oncogene[Davidson, et al. 2014]
BGN	Xq28	32	67% (12/18)	S	
GABRE	Xq28	9	29% (2/7)	S	
TAZ	Xq28	8	50% (2/4)	S	Oncogene[Chen, Xie, Huang, & Yang 2016; Moroishi, Hansen, & Guan 2015]
VMA21	Xq28	11	11% (1/9)	S	

[a] as denoted by Ensembl GRCh38

[b] Tumors are defined as escaping XCI if PPjg>0.90, subject to XCI if PPjg<0.10, or otherwise were not called.

[c] 'Mostly subject' consensus state defined by previous literature suggests these genes are subject to XCI in most tissues

[d] 'Subject' consensus state defined by previous literature suggests these genes are subject to XCI in almost all tissues

Table IV

Summary of XCI analysis results for genes thought to escape XCI with strong evidence of undergoing XCI (PP<0.10) from XCI in at least one ovarian tumor sample.

Gene	Cytoband (Ensembl)[a]	Total number of tumors evaluated	Escaping XCI (PP_{jg}>0.5)	Subject to XCI (PP_{jg}<0.5)	Proportion of samples undergoing XCI (N undergoing/N total)[b]	Consensus State[Balaton, Cotton, & Brown 2015]	Known Oncogene or Tumor Suppressor (reference)
ARSD	Xp22.33	25	11	3	21% (3/14)	Mostly E[c]	
KAL1	Xp22.31	14	1	6	86% (6/7)	Mostly E	Tumor suppressor[Liu, Cao, Chen, Xu, & Zhang 2015; Tanaka, et al. 2015]
NLGN4X	Xp22.31	19	8	2	20% (2/10)	Mostly E	
CTPS2	Xp22.2	9	0	9	100% (9/9)	Mostly E	Platinum sensitivity[Gamazon, et al. 2011]
GEMIN8	Xp22.2	11	1	4	80% (4/5)	Mostly E	
GPM6B	Xp22.2	12	3	7	70% (7/10)	Mostly E	
OFD1	Xp22.2	4	0	1	100% (1/1)	Mostly E	
RBBP7	Xp22.2	2	0	2	100% (2/2)	Mostly E	Tumor suppressor[Li, et al. 2016] [Chaligne, et al. 2015]
TCEANC	Xp22.2	9	0	7	100% (7/7)	Mostly E	Tumor suppressor with BRCA1[Hill, et al. 2014]
TRAPPC2	Xp22.2	23	10	3	23% (3/13)	Mostly E	Tumor suppressor[Tokunaga, et al. 2013]
TXLNG	Xp22.2	16	8	1	11% (1/9)	E[d]	
EIF1AX	Xp22.12	21	4	11	73% (11/15)	E	
EIF2S3	Xp22.11	18	1	4	80% (4/5)	E	
CXorf38	Xp11.4	12	1	8	89% (8/9)	E	Tumor suppressor[Bol, Xie, & Raman 2015; Dunford, et al. 2016]
DDX3X	Xp11.4	17	9	1	10% (1/10)	E	
USP9X	Xp11.4	23	3	6	67% (6/9)	Mostly E	Tumor suppressor[Jo, Kim, Yoo, & Lee 2016; Perez-Mancera, et al. 2012]
IQSEC2	Xp11.22	3	1	2	67% (2/3)	Mostly E	
KDM5C	Xp11.22	8	0	8	100% (8/8)	Mostly E	Tumor suppressor[Dalgliesh, et al. 2010; Dunford, et al. 2016]
RPS4X	Xq13.1	13	10	1	9% (1/11)	Mostly E	Prognostic marker[Tsofack, et al. 2013]
NAPIL3	Xq21.32	3	0	2	100% (2/2)	Mostly E	

[a] as denoted by Ensembl GRCh38

[b] Tumors are defined as escaping XCI if PPjg>0.90, subject to XCI if PPjg<0.10, or otherwise were not called.

[c] Consensus state defined by previous literature suggests these genes escape XCI in almost all tissues

[d] Consensus state defined by previous literature suggests these genes escape XCI in most tissue