nature genetics

Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias

Andrew Dunford[1,6], David M Weinstock[1,2,6], Virginia Savova[3,4], Steven E Schumacher[1,3], John P Cleary[2], Akinori Yoda[2], Timothy J Sullivan[1], Julian M Hess[1], Alexander A Gimelbrant[1,3,4], Rameen Beroukhim[1–3], Michael S Lawrence[5,7], Gad Getz[1,5,7] & Andrew A Lane[1,2,7]

There is a striking and unexplained male predominance across many cancer types. A subset of X-chromosome genes can escape X-inactivation, which would protect females from complete functional loss by a single mutation. To identify putative 'escape from X-inactivation tumor-suppressor' (EXITS) genes, we examined somatic alterations from >4,100 cancers across 21 tumor types for sex bias. Six of 783 non-pseudoautosomal region (PAR) X-chromosome genes (*ATRX*, *CNKSR2*, *DDX3X*, *KDM5C*, *KDM6A*, and *MAGEC3*) harbored loss-of-function mutations more frequently in males (based on a false discovery rate < 0.1), in comparison to zero of 18,055 autosomal and PAR genes (Fisher's exact $P < 0.0001$). Male-biased mutations in genes that escape X-inactivation were observed in combined analysis across many cancers and in several individual tumor types, suggesting a generalized phenomenon. We conclude that biallelic expression of EXITS genes in females explains a portion of the reduced cancer incidence in females as compared to males across a variety of tumor types.

On the basis of SEER (Surveillance, Epidemiology, and End Results Program) data from 2008–2012, US males carry an age-adjusted excess risk of 20.4% of developing any cancer (516.6 versus 411.2 for females per 100,000 person-years) and there is ≥2:1 male predominance for some individual cancer types[1]. This excess risk results in approximately 153,000 new additional cases of cancer in US men annually. Yet, the male predominance in cancer incidence remains largely unexplained[2]. The disparities between men and women occur across the world, even after adjusting for differences in gross domestic product, geographical region, and environmental risk factors, including tobacco exposure[1,3]. In fact, changes in tobacco use among males and females over the past two decades have resulted in a marked reduction in the male:female (M:F) ratio of lung and bronchus cancer (**Fig. 1a**). However, over the same time period, the M:F ratios for several cancers have remained >2:1, including for those associated with tobacco use such as kidney and renal pelvis, urinary bladder, and oral cavity and pharynx cancers (**Fig. 1a** and **Supplementary Fig. 1**).

Previous reports have identified X-chromosome genes outside of the PAR with higher mutation frequencies in male cancers than in female cancers[4–6]. For example, *KDM6A* at Xp11.2, which encodes the histone lysine demethylase UTX, has loss-of-function mutations more predominately in male cancers across multiple subtypes[4,7]. Among female cancers with *KDM6A* mutations, homozygous mutations are common[7], suggesting that some tumor-suppressor genes (TSGs) on the X chromosome outside of the PAR follow the classical Knudson two-hit hypothesis in females[8]. This poses a conundrum because one X chromosome in all female cells undergoes X-inactivation during embryogenesis, which should leave each female cell functionally haploid for non-PAR genes on the X chromosome[9]. However, a fraction of X-chromosome genes 'escape' inactivation and have biallelic expression, albeit with poorly understood mechanisms, leading to differences across individuals and cell types[10–14].

We hypothesized that mutations in TSGs that escape X-inactivation could underlie a substantial fraction of excess male cancers, as males would require only a single deleterious mutation whereas females would require two (**Fig. 1b**,**c**). We termed these genes EXITS, for 'escape from X-inactivation tumor suppressors'. In female cancers with a mutation in an EXITS gene, a corollary of the EXITS hypothesis is that the other allele will be mutated or deleted (**Fig. 1c**). The X chromosome is among the most frequently aneuploid chromosomes in female cancers[15], and, in mice, alteration of X-inactivation by targeted deletion of *Xist* promotes tumorigenesis[16].

Similarly, Y-chromosome loss is observed in ≥30% of male renal cell, head and neck, and bladder cancers[15,17,18]. Age-related loss of the Y chromosome in non-malignant blood cells is frequent, occurs at an increased rate among tobacco users, and is associated with an approximately 3.5-fold-higher risk of developing a non-hematological cancer[19,20]. The Y chromosome harbors evolutionarily ancestral homologs of a small fraction of non-PAR X-chromosome genes. X-chromosome genes with Y-chromosome homologs are known to be more likely to escape X-inactivation[21,22]. In some cases, Y-chromosome genes can rescue viability or other phenotypes upon

[1]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [2]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA. [3]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [4]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [5]Department of Pathology and Cancer Center, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [6]These authors contributed equally to this work. [7]These authors jointly directed this work. Correspondence should be addressed to A.A.L. (andrew_lane@dfci.harvard.edu).

Received 5 April; accepted 24 October; published online 21 November 2016; doi:10.1038/ng.3726

loss of their X-chromosome homologs[23,24], although for most genes there is not clear evidence of functional redundancy. Thus, a second corollary to the EXITS hypothesis is that Y-chromosome loss will co-occur in tumors in males that have a mutation in an EXITS gene with a functional Y-chromosome homolog (**Fig. 1c**).

RESULTS

Male-biased loss-of-function mutations on the X chromosome in cancer

To test the EXITS hypothesis, we performed an unbiased analysis of paired tumor–germline exome sequencing data from 4,126 patients across 21 tumor types from The Cancer Genome Atlas (TCGA) and Broad Institute data sets (**Supplementary Fig. 2** and **Supplementary Table 1**). In 1,994 cases, copy number variation data were also available based on high-density SNP arrays. The tumors analyzed excluded sex-restricted cancers such as prostate and ovarian cancers, as well as breast cancer.

We used a conservative mutation classification algorithm to ensure that the included variants were restricted to somatic truncating and missense alterations that were most likely loss of function (Online Methods) or DNA copy number loss of at least one allele (**Supplementary Fig. 3**). We applied a permutation analysis to determine which genes were mutated in males at frequencies higher than expected on the basis of the overall mutation rates in males and females for each tumor type and normalized to the number of X chromosomes. We performed this analysis for only loss-of-function mutations (including likely deleterious single-nucleotide variants (SNVs) and indels) in the 4,126 patients with exome data and then again for loss-of-function mutation or copy number loss in the 1,994 patients with both exome and copy number data available.

At false discovery rate (FDR) < 0.1, there were no autosomal or PAR genes (n = 18,055) that had loss-of-function mutations more frequently in male cancers. In addition, at FDR < 0.1, no genes on the X chromosome had a significantly increased frequency of silent coding mutations in male cancers, and no genes on the X chromosome had non-silent loss-of-function mutations more frequently in female cancers (**Supplementary Tables 2–4**). In contrast, at FDR < 0.1, 6 of 783 non-PAR X-chromosome genes (*ATRX*, *CNKSR2*, *DDX3X*, *KDM5C*, *KDM6A*, and *MAGEC3*) harbored loss-of-function mutation or loss-of-function mutation/copy number loss more frequently in male cancers (Fisher's exact $P < 0.0001$ as compared to zero of 18,055 autosomal and PAR genes; **Fig. 2a**,**b**, **Table 1**, and **Supplementary Tables 5** and **6**). Of note, *ATRX*[25,26], *DDX3X*[27], *KDM5C*[28], and *KDM6A*[29] have previously been implicated as TSGs via recurrent loss-of-function alterations in cancer genomes and/or by direct experimental evidence.

We performed the same permutation analysis in each of the 21 cancer types individually to discover tissue- or disease-restricted EXITS genes. Loss-of-function mutation or copy number loss was enriched in *ATRX* among male lower-grade gliomas (LGG; FDR < 10^{-4}) and in *KDM5C* among male clear cell kidney cancers (KIRC; FDR = 0.044) (**Fig. 2c**,**d** and **Table 1**). There were no autosomal or PAR genes with male-biased loss-of-function mutation in any of the individual cancer types at FDR < 0.1.

To assess the robustness of these findings, we then used a second statistical test on the same data set based on a log-likelihood ratio, which also normalizes to the background male and female mutation rates in each tumor type and the number of X chromosomes (Online Methods). Five of the six genes (*CNKSR2*, *DDX3X*, *KDM5C*, *KDM6A*, and *MAGEC3*) were rediscovered as significantly more frequently mutated across all male cancers in this analysis (FDR < 0.1), as were *ATRX* in male LGG and *KDM5C* in male KIRC (**Supplementary Fig. 4** and **Supplementary Table 7**).

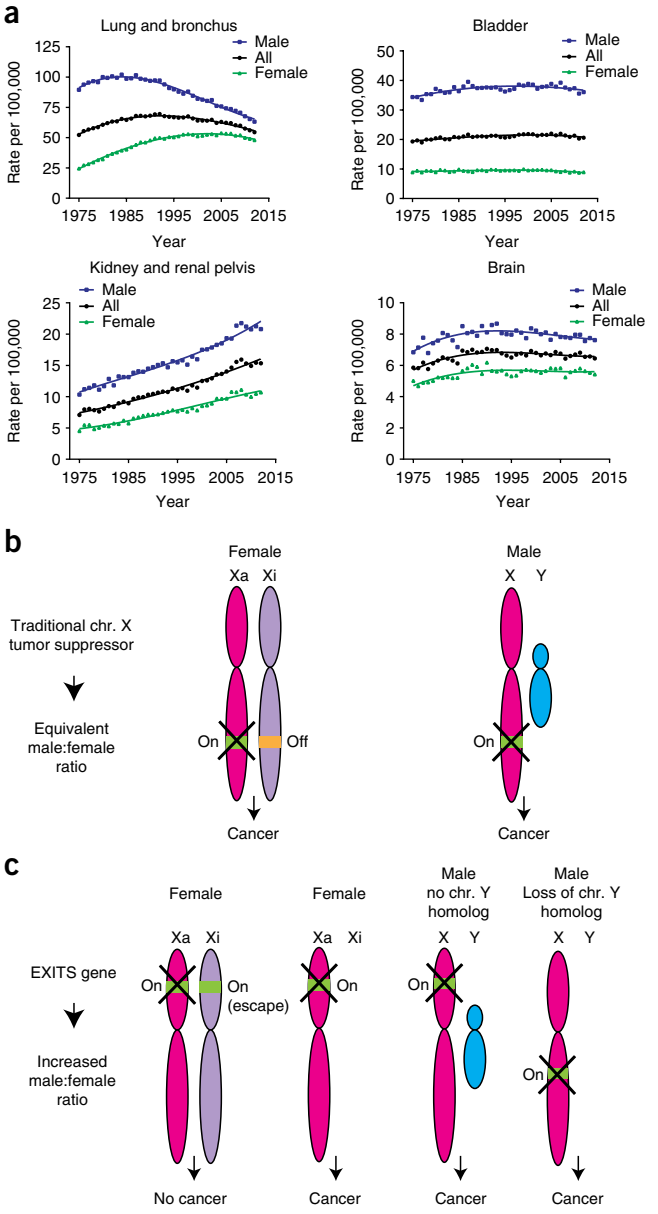

Figure 1 Escape from X-inactivation tumor-suppressor (EXITS) genes. (**a**) SEER data of annual incidence rates over time for the indicated cancer types in males (blue), females (green), or all patients (black). (**b**,**c**) The EXITS hypothesis. (**b**) 'Traditional' tumor-suppressor genes (TSGs) on the X chromosome for females and males. A single deleterious mutation in a TSG is equally likely to occur in male and female cancers because males have only one X chromosome and females have one active X chromosome (Xa; pink) and one inactive X chromosome (Xi; purple). (**c**) Model for EXITS gene behavior. In females, there are two active alleles of EXITS genes, and females are therefore protected from complete gene loss after a single alteration. Complete inactivation of an EXITS gene may require biallelic mutations, or mutation with loss of the other X chromosome. In males, one mutation could inactivate the only allele of an EXITS gene that has no functional Y-chromosome homolog, and males would therefore be more likely to develop cancers associated with mutations in those TSGs. Alternatively, because some genes that escape X-inactivation have Y-chromosome homologs with redundant function, cancers with mutations in those genes would be more likely to occur in males who also have somatic loss of the Y chromosome.

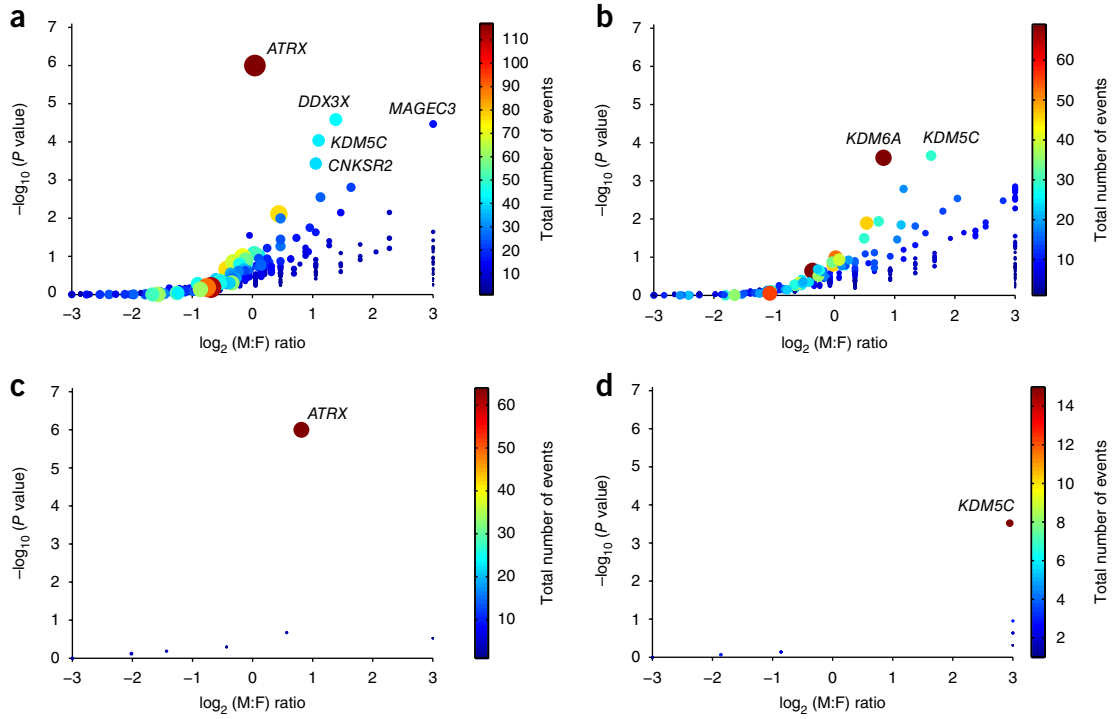

Figure 2 Genes with higher frequencies of somatic loss-of-function alterations in male cancers. (**a**,**b**) Permutation testing for genes on the X chromosome across all cancer data sets is shown. The $\log_2$ (M:F) ratio of events is plotted for each gene against the significance (*P*) value. The size and color of each circle correspond to the number of loss-of-function mutations (**a**) or loss-of-function mutations and copy number loss events (**b**) in that gene. Genes with significantly higher (FDR < 0.1) frequencies of mutation in male cancers are labeled. (**c**,**d**) Disease-specific permutation testing of loss-of-function mutations in lower-grade glioma (LGG) (**c**) and clear cell kidney cancer (KIRC) (**d**).

EXITS gene alterations associated with excess male cancers

Adult LGGs occur at an M:F ratio of ~1.4:1 (or 58.3% male:41.7% female)[30]. Considering that 42% of male LGGs in our data set had *ATRX* mutations as compared to 26% of female LGGs, approximately 80% of the excess male LGGs in our cohort had *ATRX* mutations. Similarly, 8.6% of the excess male head and neck squamous cell carcinomas (HNSC) had *KDM6A* mutations and 12.1% of the excess male KIRC cases had *KDM5C* mutations (see the Online Methods for calculations). However, these figures are undoubtedly an underestimate of the total contribution of male-biased alterations in these diseases because they only consider stringently defined, loss-of-function mutations within coding regions and they are limited by the sample size of our data sets.

To address the latter limitation, we calculated the number of tumor–normal pairs needed for 80% power to detect a coding loss-of-function mutation on the X chromosome with a significant male bias given different M:F disease incidence ratios, basal mutation rates, and prevalence of the alteration in a specific population (**Fig. 3a**). For example, KIRC has an M:F incidence ratio of ~2:1 and, therefore, as expected, ~50% of all mutations on the X chromosome across all KIRC cases are in males because they have only one X chromosome. If an EXITS gene were mutated in 5% of all KIRC cases, >1,000 tumors would need to be sequenced to have 80% power to detect a fourfold male mutation bias. Therefore, many additional cancers, including hundreds of tumor types not represented in this data set, will need to be assessed before the contribution of EXITS genes to overall excess

Table 1 Genes with significantly (FDR < 0.1) increased M:F mutation ratios identified by permutation analysis

Gene	Analysis set	LOF mutations	Total cancers	*P* value	*Q* (FDR) value
ATRX	All	70 M: 47 F	2,440 M: 1,686 F	0.000001	0.00066
ATRX	LGG	45 M: 19 F	98 M: 72 F	0.000001	0.000071
CNKSR2	All	30 M: 10 F	2,440 M: 1,686 F	0.00037	0.049
DDX3X	All	34 M: 9 F	2,440 M: 1,686 F	0.000026	0.0075
KDM5C	All	31 M: 10 F	2,440 M: 1,686 F	0.000092	0.015
KDM5C	KIRC	14 M: 1 F	216 M: 118 F	0.0003	0.044
MAGEC3	All	15 M: 1 F	2,440 M: 1,686 F	0.000034	0.0075
Gene	Analysis set	LOF mutations or CN deletions	Total cancers	*P* value	*Q* (FDR) value
KDM5C	All	24 M: 5 F	1,225 M: 769 F	0.00022	0.079
KDM5C	KIRC	14 M: 1 F	216 M: 118 F	0.00047	0.08
KDM6A	All	50 M: 18 F	1,225 M: 769 F	0.00025	0.079

Significance values are based on deviation of the observed mutation incidence in a specific gene relative to that expected in a given set. This approach normalizes to the number of male and female cancers (and to the number of X chromosomes) as well as to the background mutation incidence in male and female cancers in a given set. LGG, lower-grade glioma; KIRC, clear cell kidney cancer; all, pooled data from all included cancer types; LOF, loss of function (Online Methods); CN, copy number; FDR, false discovery rate.

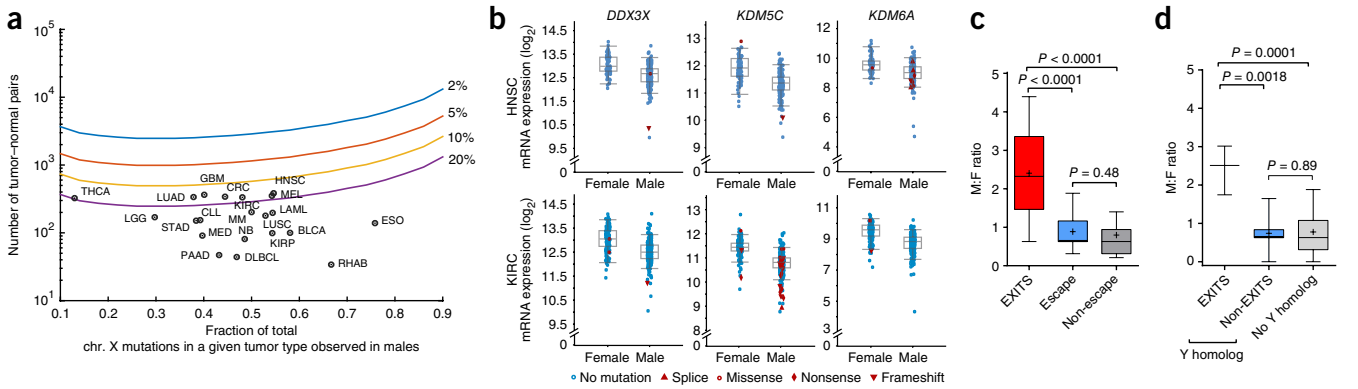

Figure 3 EXITS gene alterations are associated with male cancers. (**a**) Calculation of the number of tumor–normal pairs needed for 80% power to detect fourfold male-biased loss-of-function mutations with Bonferroni-corrected $P < 0.1$ (that is, mutations with four times greater prevalence in tumors from males as compared to females). The *x* axis represents the fraction of all mutations on the X chromosome occurring in males in the cohort (a function of the M:F ratio of disease incidence and overall mutation rate in males and females). Lines represent the percentage of cancers in a given tumor type that harbor a specific mutation (blue, 2%; red, 5%; yellow, 10%; purple, 20%). Each of the 21 tumor types we analyzed is plotted to show the power we had to detect a male-biased mutation on the basis of the fraction of mutations on the X chromosome in males and the number of tumor–normal pairs in the data set. BLCA, bladder carcinoma; CLL, chronic lymphocytic leukemia; CRC, colorectal carcinoma; DLBCL, diffuse large B cell lymphoma; ESO, esophageal carcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KIRC, clear cell kidney cancer; KIRP, papillary kidney cancer; LAML, acute myeloid leukemia; LGG, lower-grade glioma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MED, medulloblastoma; MEL, melanoma; MM, multiple myeloma; NB, neuroblastoma; PAAD, pancreatic ductal adenocarcinoma; RHAB, rhabdoid tumor; STAD, stomach adenocarcinoma; THCA, thyroid carcinoma. (**b**) RNA–seq expression levels ($\log_2$) for *DDX3X*, *KDM5C*, and *KDM6A* in head and neck squamous cell carcinoma (HNSC) and clear cell kidney cancer (KIRC) in the TCGA data sets, with tumors separated by patient sex (data visualization from http://www.cbioportal.org/). Each dot represents one tumor: blue symbols are tumors with no mutation in the gene, and red symbols are tumors with a mutation of the indicated type ($P < 0.0001$ for all female–male expression comparisons by Kolmogorov–Smirnov test, either including or excluding mutated cases; see also **Supplementary Fig. 8**). Bar, median; box, interquartile range; whiskers, 10th–90th percentiles. Number of samples: HNSC, 203 male, 76 female; KIRC, 269 male, 144 female. (**c**) M:F ratios of loss-of-function mutations in the EXITS genes identified in **Table 1**, all other X-chromosome escape genes ($n = 56$), and X-chromosome non-escape genes (data compared by *t* test; bar, median; plus sign, mean; box, interquartile range; whiskers,10th–90th percentile). (**d**) M:F ratios of loss-of-function mutations in the EXITS genes that have functional Y-chromosome homologs (*DDX3X*, *KDM5C*, and *KDM6A*), all other X-chromosome genes with predicted functional Y-chromosome homologs ($n = 14$), and X-chromosome genes without a Y-chromosome homolog (data compared by *t* test; plotted as in **c**).

male cancer risk can be fully quantified. This calculation provides target sample sizes for such studies.

The expression of EXITS genes could also be downregulated in the absence of coding mutations by several mechanisms, including noncoding mutations or epigenetic changes. We analyzed exome and RNA–seq data by patient sex for HNSC and KIRC cases. Essentially all tumors with downregulation of EXITS genes (defined by expression below the 5th percentile of expression in male cancers) in the absence of a DNA mutation were from males (22/469 males versus 2/218 females for *DDX3X*, $P = 0.012$ by Fisher's exact test; 16/450 males versus 2/216 females for *KDM5C*, $P = 0.071$; 23/465 males versus 2/217 females for *KDM6A*, $P = 0.0077$; **Fig. 3b**). Whole-genome sequencing of a large number of tumors will be necessary to determine whether noncoding mutations on the X chromosome underlie EXITS gene downregulation.

EXITS gene mutations associated with loss of the paired Y or X chromosome

Next, we asked whether additional genes that escape X-inactivation or have Y-chromosome homologs might also function as EXITS genes. We compiled a list of 59 X-chromosome genes with evidence of escape from X-inactivation in multiple contexts from studies of human lymphoblastoid cells and hybrid fibroblasts[12,13,31] (**Supplementary Table 8**). These 59 genes had a higher M:F ratio of loss-of-function mutation or copy number loss than other X-chromosome genes ($P = 0.022$; **Supplementary Fig. 5a**). Similarly, we identified 17 X-chromosome genes with predicted functional Y-chromosome homologs[21] (**Supplementary Table 8**) and found that they had a trend toward a higher M:F ratio for loss-of-function mutation or copy number loss in comparison to other X-chromosome genes ($P = 0.058$; **Supplementary Fig. 5b**).

If we excluded the 6 genes we identified as EXITS genes, there was no discernible male mutation bias among either the 56 remaining escape genes or the 14 remaining X-chromosome genes with functional Y-chromosome homologs (**Fig. 3c,d**). However, we may have failed to identify additional EXITS genes from our data set because of power limitations resulting from agnostically assessing all coding genes. A permutation test limited to genes previously reported to either escape X-inactivation ($n = 59$) or have Y-chromosome homologs in humans or in recent mammalian evolution ($n = 33$)[21] identified 2 additional genes with higher frequencies of loss-of-function mutation and/or copy number loss in male cancers (FDR < 0.1): *NLGN4X* and *RBM10* (**Supplementary Table 9**). Previous functional data support *RBM10* as a bona fide TSG[32].

We next tested the hypothesis that EXITS genes will harbor biallelic inactivation when mutated in female tumors. We used SNP array data to infer which female tumors had lost an entire X chromosome and compared the frequency of this event to the incidence of loss-of-function mutation or focal copy number loss on the remaining X chromosome (**Supplementary Fig. 3**). Tumors in females with EXITS gene mutations were more likely to have lost the whole other X chromosome than tumors without EXITS gene mutations (23.3% (10/43) versus 6.2% (45/726), $P = 0.0005$ by Fisher's exact test), suggesting enrichment for biallelic loss. In addition, 6.3% of the tumors in females with an EXITS gene mutation had two loss-of-function mutations in that gene, although with short-read sequencing we were unable to determine whether these were in *cis* or *trans*.

To determine whether Y-chromosome loss is enriched among tumors in males with EXITS gene mutations, we used a conservative approach to classify Y-chromosome copy number on the basis of exome sequencing data across 1,443 tumors from males in our data set (Online Methods and **Supplementary Fig. 6**). Tumors in males with loss-of-function mutations in any of the three EXITS genes with Y-chromosome homologs (*DDX3X*, *KDM5C*, and *KDM6A*) had a trend toward enrichment for Y-chromosome loss in comparison to tumors without mutations (10.2% (9/88) versus 5.8% (78/1,355), $P = 0.077$). Tumors in males with a loss-of-function mutation in any X-chromosome gene with a predicted functional Y-chromosome homolog were more likely to have lost the Y chromosome than those without such a mutation (10.6% (15/142) versus 5.5% (72/1,301), $P = 0.019$). Therefore, age- and tobacco-associated spontaneous loss of the Y chromosome could disproportionately increase the frequency of some cancers in males, as loss-of-function mutation or copy number loss of EXITS genes with functional Y-chromosome homologs would lead to complete gene inactivation in male cells having lost the Y chromosome (**Fig. 1c**)[19].

To assess the relative functional contribution of X–X pairs in females as compared to X–Y homologs in males, we compared the rate of concurrent mutation and X-chromosome loss in tumors from females with the rate of concurrent mutation and Y-chromosome loss in tumors from males for the three EXITS genes with Y-chromosome homologs (*DDX3X*, *KDM5C*, and *KDM6A*). Tumors from females with loss-of-function or copy number mutation were more likely to lose the X chromosome than tumors from males with loss-of-function or copy number mutation were to lose the Y chromosome (36% (9/25) versus 8.2% (6/73), $P = 0.0022$). This result suggests that the Y-chromosome homologs in males may not have tumor-suppressor activity equivalent to that of the EXITS gene alleles on the inactivated X chromosome in females.

Hypothesis generation from sex-biased mutation patterns

Among the EXITS genes we identified, *DDX3X*, *KDM5C*, and *KDM6A* are recognized escape genes across multiple tissues[12,13]. *CNKSR2* and *MAGEC3* were more recently suggested to escape X-inactivation as determined by next-generation sequencing and epigenetic analyses[33]. There are data in human cell lines showing *ATRX* escape from X-inactivation using RNA FISH[34,35], but *ATRX* has not traditionally been classified among the escape genes[12,13,21,31]. The *ATRX* locus is located in an X-chromosome region that contains multiple escape genes and/or Y-chromosome homologs in lower mammals[21], and *ATRX* escapes X-inactivation in human trophoblast cells[36]. Nearly all of the excess male cancers with *ATRX* mutations in our data set were LGG cases (**Table 1**). We therefore hypothesized that *ATRX* may escape X-inactivation but only in certain contexts, possibly including in the brain. Tissue-specific and between-individual variation in X-inactivation escape are recognized phenomena that have been described previously[14,33].

We took two approaches using RNA–seq to demonstrate evidence of X-inactivation escape in putative EXITS genes in tumors and normal tissues, constituting analyses of allele-specific expression in female cells and male versus female expression levels. At germline heterozygous SNP sites in coding regions from tumors without somatic mutations in the tested genes, we observed evidence of biallelic expression in females of *ATRX*, *KDM6A*, *KDM5C*, and *DDX3X* in one or more of six tumor types (glioblastoma multiforme (GBM), LGG, HNSC, KIRC, lung adenocarcinoma (LUAD), and lung squamous cell carcinoma (LUSC)) in which we had identified male-biased mutations (**Supplementary Fig. 7**). This included tumor types where male-biased mutations were significant in single-cancer analysis (for example, *ATRX* in LGG and *KDM5C* in KIRC). We also analyzed female and male expression levels of EXITS genes in tumors without mutations because escape from X-inactivation often results in higher expression in female cells[11]. *DDX3X*, *KDM5C*, and *KDM6A* had higher expression in non-mutated female than in male tumors across all cancer types tested. Notably, *ATRX* expression was only higher in female LGG cases and *CNKSR2* expression was higher only in female LUAD cases, the tumor types for each gene where the male loss-of-function mutation bias was also seen (**Supplementary Fig. 8**).

To analyze biallelic expression in normal tissues, we queried data from the Genotype-Tissue Expression (GTEx) Project, which includes RNA–seq data from multiple tissue types in non-diseased individuals[37]. Within local tissue environments, X-inactivation can be skewed toward one chromosome as a result of a shared developmental origin[38], which might allow detection of escape from X-inactivation if the GTEx biopsy demonstrated higher minor allele expression for escape genes than for 'known' non-escape genes. Indeed, that is what we observed for EXITS genes, including high minor allele expression of *ATRX* in brain biopsies from several donors at multiple anatomical sites, suggesting escape from X-inactivation ($P < 0.0001$ in the cortex and $P = 0.0224$ in the cerebellum in comparison to non-escape genes by Kolmogorov–Smirnov test; **Supplementary Fig. 9**).

We also analyzed EXITS gene expression by sex in normal tissue biopsies from GTEx. In contrast to the expression pattern in other tissues and male brain, *ATRX* expression in female brain showed two distribution peaks and was best fit by models of bimodality ($P = 0.04$; **Supplementary Fig. 10a**,**b**). We compared female and male expression for the six EXITS genes across multiple tissues and found that, while most had higher expression in females at all measurable sites, evidence for *ATRX* escape from X-inactivation was limited to female brain (**Supplementary Fig. 10c**). These data also suggest that *ATRX* undergoes heterogeneous escape from X-inactivation in female brain. Further studies will be required to understand whether *ATRX* escape is restricted to certain brain cell subsets or regions, or whether there is between-female heterogeneity in escape status, and how these phenomena could contribute to male bias among *ATRX*-mutated gliomas.

CNKSR2 was not previously implicated as a TSG in large sequencing studies[39]. It encodes a putative Ras-pathway scaffolding protein implicated in an X-linked neurodevelopmental disorder[40]. Depletion of *Cnksr2* from mouse NIH 3T3 fibroblasts using two independent lentiviral short hairpin RNAs (shRNAs) resulted in expression changes that were enriched in genes associated with oncogene signatures by gene set enrichment analysis (GSEA) in the Molecular Signatures Database (MSigDB; Broad Institute) C2:CGP collection[41], including genes involved in transformation by HRAS, KRAS, and the Ras-like protein RHOA (**Supplementary Fig. 11a–c** and **Supplementary Table 10**). In addition, we queried the 50 'Hallmarks' MSigDB data sets of genes involved in general cellular function[42], a combined automated and manual curation of >10,000 gene sets, to overcome problems of redundancy and heterogeneity in GSEA. *Cnksr2* depletion was most enriched with signatures of mTOR and KRAS signaling (**Supplementary Fig. 11d** and **Supplementary Table 11**). In line with activation of the RAS–MAPK pathway, cells expressing *Cnksr2* shRNAs had increased ERK phosphorylation and enhanced colony-forming activity in soft agar, consistent with *in vitro* transformation (**Supplementary Fig. 11e–g**). These findings demonstrate how genetic–epidemiological associations such as sex bias may help distinguish driver from passenger events and possibly identify cancer-associated genes with a higher likelihood of functional relevance.

DISCUSSION

Escape from X-inactivation results in expression of two copies of a TSG on the X chromosome in females, whereas males only have one copy. The EXITS hypothesis is that biallelic expression of these genes affords females enhanced cancer protection, which substantively contributes to the observed higher incidence of some tumors in males. Our data have provided evidence to support this hypothesis, as the genes with an increased incidence of loss-of-function mutations in males across many cancer types were exclusively non-PAR genes on the X chromosome, including several that are known to be TSGs and/or known to escape X-inactivation in certain contexts. Genes that do not normally escape X-inactivation could also function as EXITS genes in some situations, as it is recognized that aberrant or 'leaky' X-inactivation or escape can occur in cancer cells[35,43]. Additional studies are needed to identify the full complement of sex-biased cancer-related genes; these will include the sequencing of more tumor–normal pairs to increase statistical power, assessments of noncoding genomes for additional mechanisms of EXITS gene inactivation, and interrogation of additional tumor types not represented herein to identify disease-restricted sex-biased TSGs.

Undoubtedly, EXITS gene mutations are not the sole explanation for differences in cancer incidence between men and women. Alcohol use, tobacco exposure, and endocrine biology are known to affect cancer epidemiology, with the latter possibly causing female predominance of some cancers. However, these environmental and hormonal factors associated with sex-specific differences in cancer could interact with EXITS loci or their gene products to modulate cancer risk. For example, a clastogenic carcinogen could disproportionately affect men if the process of carcinogenesis involves inactivation of an EXITS gene or loss of the Y chromosome, as has been shown to result from tobacco smoking[19]. Similarly, enhancers or other regulatory elements on the X chromosome that modulate autosomal gene transcription in *trans* could qualify as 'noncoding EXITS' if their loss promotes carcinogenesis.

Another implication of our data is that tumors in males and females with the same cancer type may have distinct genetics, separate from the number of X and Y chromosomes. These differences could be directly related to the impact of EXITS gene mutations on disease biology or drug response. Alternatively, they could result in more general effects in a population because male and female cancers of a specific type may be effectively distinct diseases (that is, tumors from males without EXITS gene mutations might behave similarly to tumors from females, while EXITS-gene-mutated cancers could have their own biology). To begin to explore these possibilities, we propose that clinical oncology studies should be statistically powered to understand sex-specific differences in outcomes resulting from distinct tumor genetics. In addition, preclinical models can be generated to address the relative contributions of cell-intrinsic and cell-extrinsic differences between male and female cancers.

Finally, we note that several of the EXITS genes identified herein have also been implicated in germline genetic neurodevelopmental diseases[40,44–47] and that non-malignant disorders with a significantly higher risk among males (for example, autism[48] and schizophrenia[49]) have been associated with polymorphisms on the X chromosome. These data suggest that, besides the obvious link between male sex and monogenic diseases associated with mutation of X-chromosome genes, the haploid nature of the X chromosome in males may increase the risk of developing polygenic diseases other than cancer.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Accession codes. RNA–seq data are available through the Gene Expression Omnibus (GEO) under accession GSE85462.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank C. Sievers and B. Bernstein for helpful discussions, Z. Herbert and the Molecular Biology Core Facilities at the Dana-Farber Cancer Institute for assistance with RNA–seq, and T. Golub for facilitating the project. This work was supported by National Cancer Institute grant K08CA181340 (A.A.L.), an American Society of Hematology Scholar Award (A.A.L.), and a V Foundation Scholar Award (A.A.L.). G.G. is the Paul C. Zamecnik, MD, Chair in Oncology at Massachusetts General Hospital. D.M.W. is a Leukemia and Lymphoma Society Scholar and is supported by a Stand Up to Cancer Innovative Research Grant.

AUTHOR CONTRIBUTIONS

A.D., V.S., S.E.S., J.P.C., A.Y., T.J.S., and J.M.H. designed, performed, and analyzed the data from computational and laboratory analyses. D.M.W., A.A.G., R.B., M.S.L., G.G., and A.A.L. conceived the study, designed experiments, and interpreted the data. A.D., D.M.W., M.S.L., G.G., and A.A.L. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Edgren, G., Liang, L., Adami, H.O. & Chang, E.T. Enigmatic sex disparities in cancer incidence. *Eur. J. Epidemiol.* **27**, 187–196 (2012).
2. Cook, M.B. *et al.* Sex disparities in cancer incidence by period and age. *Cancer Epidemiol. Biomarkers Prev.* **18**, 1174–1182 (2009).
3. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. *CA Cancer J. Clin.* **63**, 11–30 (2013).
4. Van der Meulen, J. *et al.* The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. *Blood* **125**, 13–21 (2015).
5. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
6. Van Vlierberghe, P. *et al. PHF6* mutations in T-cell acute lymphoblastic leukemia. *Nat. Genet.* **42**, 338–342 (2010).
7. van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene *UTX* in human cancer. *Nat. Genet.* **41**, 521–523 (2009).
8. Knudson, A.G. Antioncogenes and human cancer. *Proc. Natl. Acad. Sci. USA* **90**, 10914–10921 (1993).
9. Yang, C. *et al.* X-chromosome inactivation: molecular mechanisms from the human perspective. *Hum. Genet.* **130**, 175–185 (2011).
10. Berletch, J.B. *et al.* Escape from X inactivation varies in mouse tissues. *PLoS Genet.* **11**, e1005079 (2015).
11. Berletch, J.B., Yang, F., Xu, J., Carrel, L. & Disteche, C.M. Genes that escape from X inactivation. *Hum. Genet.* **130**, 237–245 (2011).
12. Carrel, L. & Willard, H.F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400–404 (2005).
13. Johnston, C.M. *et al.* Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. *PLoS Genet.* **4**, e9 (2008).
14. Talebizadeh, Z., Simon, S.D. & Butler, M.G. X chromosome gene expression in human tissues: male and female comparisons. *Genomics* **88**, 675–681 (2006).
15. Duijf, P.H., Schultz, N. & Benezra, R. Cancer cells preferentially lose small chromosomes. *Int. J. Cancer* **132**, 2316–2326 (2013).
16. Yildirim, E. *et al. Xist* RNA is a potent suppressor of hematologic cancer in mice. *Cell* **152**, 727–742 (2013).
17. Veiga, L.C., Bérgamo, N.A., Reis, P.P., Kowalski, L.P. & Rogatto, S.R. Loss of Y-chromosome does not correlate with age at onset of head and neck carcinoma: a case–control study. *Braz. J. Med. Biol. Res.* **45**, 172–178 (2012).
18. Zhang, L.J., Shin, E.S., Yu, Z.X. & Li, S.B. Molecular genetic evidence of Y chromosome loss in male patients with hematological disorders. *Chin. Med. J. (Engl.)* **120**, 2002–2005 (2007).
19. Dumanski, J.P. *et al.* Smoking is associated with mosaic loss of chromosome Y. *Science* **347**, 81–83 (2015).

20. Forsberg, L.A. *et al.* Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat. Genet.* **46**, 624–628 (2014).
21. Bellott, D.W. *et al.* Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014).
22. Cortez, D. *et al.* Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488–493 (2014).
23. Shpargel, K.B., Sengoku, T., Yokoyama, S. & Magnuson, T. UTX and UTY demonstrate histone demethylase–independent function in mouse embryonic development. *PLoS Genet.* **8**, e1002964 (2012).
24. Welstead, G.G. *et al.* X-linked H3K27me3 demethylase Utx is required for embryonic development in a sex-specific manner. *Proc. Natl. Acad. Sci. USA* **109**, 13004–13009 (2012).
25. Brat, D.J. *et al.* Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
26. Pinto, E.M. *et al.* Genomic landscape of paediatric adrenocortical tumours. *Nat. Commun.* **6**, 6302 (2015).
27. Jiang, L. *et al.* Exome sequencing identifies somatic mutations of *DDX3X* in natural killer/T-cell lymphoma. *Nat. Genet.* **47**, 1061–1066 (2015).
28. Dalgliesh, G.L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360–363 (2010).
29. Ntziachristos, P. *et al.* Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia. *Nature* **514**, 513–517 (2014).
30. Dubrow, R. & Darefsky, A.S. Demographic variation in incidence of adult glioma by subtype, United States, 1992–2007. *BMC Cancer* **11**, 325 (2011).
31. Wilson Sayres, M.A. & Makova, K.D. Gene survival and death on the human Y chromosome. *Mol. Biol. Evol.* **30**, 781–787 (2013).
32. Bechara, E.G., Sebestyén, E., Bernardis, I., Eyras, E. & Valcárcel, J. RBM5, 6, and 10 differentially regulate *NUMB* alternative splicing to control cancer cell proliferation. *Mol. Cell* **52**, 720–733 (2013).
33. Cotton, A.M. *et al.* Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol.* **14**, R122 (2013).
34. Al Nadaf, S. *et al.* A cross-species comparison of escape from X inactivation in Eutheria: implications for evolution of X chromosome inactivation. *Chromosoma* **121**, 71–78 (2012).
35. Chaligné, R. *et al.* The inactive X chromosome is epigenetically unstable and transcriptionally labile in breast cancer. *Genome Res.* **25**, 488–503 (2015).
36. Patrat, C. *et al.* Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. *Proc. Natl. Acad. Sci. USA* **106**, 5198–5203 (2009).
37. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
38. Wu, H. *et al.* Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. *Neuron* **81**, 103–119 (2014).
39. Lawrence, M.S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
40. Vaags, A.K. *et al.* Absent CNKSR2 causes seizures and intellectual, attention, and language deficits. *Ann. Neurol.* **76**, 758–764 (2014).
41. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
42. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
43. Chaligné, R. & Heard, E. X-chromosome inactivation in development and cancer. *FEBS Lett.* **588**, 2514–2522 (2014).
44. Gibbons, R.J. & Higgs, D.R. Molecular-clinical spectrum of the ATR-X syndrome. *Am. J. Med. Genet.* **97**, 204–212 (2000).
45. Jensen, L.R. *et al.* Mutations in the *JARID1C* gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am. J. Hum. Genet.* **76**, 227–236 (2005).
46. Miyake, N. *et al. MLL2* and *KDM6A* mutations in patients with Kabuki syndrome. *Am. J. Med. Genet. A.* **161A**, 2234–2243 (2013).
47. Snijders Blok, L. *et al.* Mutations in *DDX3X* are a common cause of unexplained intellectual disability with gender-specific effects on Wnt signaling. *Am. J. Hum. Genet.* **97**, 343–352 (2015).
48. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
49. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

ONLINE METHODS

SEER data. SEER data were downloaded from http://seer.cancer.gov/data/.

Tumor–normal exon sequencing, DNA copy number, and RNA sequencing data. Genomic data were obtained from the TCGA data portal (https://tcga-data.nci.nih.gov/tcga/) and from additional Broad Institute data sets[39,50]. Informed consent was obtained from all subjects by local institutional review boards, and consents were reviewed by the sequencing centers, according to TCGA guidelines. To visualize gene expression by mutation status for males and females, TCGA data were downloaded and visualized using cBioPortal (http://www.cbioportal.org/)[51,52].

Mutation classification. Tumor-associated DNA variants were called 'truncating' if they resulted in nonsense, insertion, deletion, translation start site, or non-stop mutations. Truncating and likely functional missense (loss-of-function) mutations included all truncating mutations as well as missense mutations that occurred at the same site in at least three patients (n = 419 across the pan-cancer data set, of which 9 were male predominant) or had a Cons 46 Vertebrates score (http://ucscbrowser.genap.ca/cgi-bin/hgTrackUi?db=hg19&g=cons46way)[53] of greater than 0.9, indicating high conservation across species (n = 3,083, of which 124 were male predominant).

Permutation analysis. For each set of events, the probability of observing the number of events seen in males was determined by a series of permutations. Within each tumor type, the probability that any given gene alteration would occur in a male was assumed to be directly related to the number of mutations across all males, divided by the total number of mutations across all samples. This controlled for the number of males in the set as well as the relative mutation rate on the X chromosome in males versus females in the set. For instance, if 60% of the coding mutations on the X chromosome in the data set for a tumor type were in males, then for any given X-chromosome gene, assuming there is no selective advantage for males versus females, we should see ~60% of the events occurring in males. For each gene, the total number of events was counted and a set of random permutations was constructed in which each event was randomly assigned a 'male' or 'female' value based on this calculated probability. For the pan-cancer analyses, the probability of an event being in a male was calculated on a per-tumor-type basis, and events were counted and permuted within each tumor type. Once the number of males in the permutation was calculated, it was compared against the observed data to see whether the number of males randomly generated in this manner was greater than or equal to that observed in the data. For each gene, 1 million permutations were carried out, and the fraction of permutations with a greater than or equal number of male events was reported as the P value. Multiple-hypothesis correction was then performed on the complete set of genes that were affected by the set of events examined, and genes with FDR <0.1 (by Benjamini–Hochberg FDR correction) were reported as significant. For permutations on truncating and loss-of-function mutations, this probability was calculated on the basis of the coding mutation count for the X chromosome. For permutations on copy-loss events, the frequency of copy-loss events on the X chromosome was used. For the combined loss-of-function mutation/copy number loss permutations, the sum of these two was used.

Log-likelihood ratio test. Assuming that the only factor differentiating the male mutation rate of a gene and the female mutation rate of a gene is the difference in the background mutation rate on the X chromosome (that is, assuming that there is no selective bias in males for mutating a gene), then the probability of a male having a mutation in a gene should be directly related to the probability of a female having a mutation in a gene, corrected for the number of copies of the X chromosome in females (n = 2) and males (n = 1)

$$p_{\text{female}} = r \times p_{\text{male}}$$

where r is the F:M ratio of coding mutations across the X chromosome; if females in a set have twice the number of coding mutations on the X chromosome as males, it can be expected that any given gene on the X chromosome will be twice as likely to be mutated in a female patient than in a male patient. For each test, the actual F:M ratio of coding mutations on the X chromosome in the analysis set was calculated and used. Because there is a direct relationship between the probability of a mutation in males in this model and a mutation in females, we can express the maximum likelihood of the observed data given this null model, by maximizing it with respect to the single value of P,

$$L_0 = \max((p_{\text{male}})^m (1-p_{\text{male}})^{M-m} (r \times p_{\text{male}})^f (1-r \times p_{\text{male}})^{F-f})$$

where M and F are the total number of male and female patients, respectively, and m and f are the number of mutated males and females. The alternative hypothesis is that there are independent factors affecting males and females, in which case two values of p are needed to calculate the likelihood that the data fit the model

$$L_1 = \max((p_{\text{male}})^m (1-p_{\text{male}})^{M-m} (p_{\text{female}})^f (1-p_{\text{female}})^{F-f})$$

which can be maximized by using the observed mutation counts

$$L_1 = (m/M)^m (M-m/M)^{M-m} (f/F)^f (F-f/F)^{F-f}$$

The log-likelihood ratio (LLR) is calculated simply by taking the log of the ratio of these two numbers

$$\text{LLR} = \log(L_1 / L_0)$$

which, using Wilks' theorem, was converted to a P value for each gene. The Benjamini–Hochberg procedure for controlling the FDR was then applied, and genes with FDR <0.1 were reported as significant.

Copy number determination by SNP array. Probe-level signal intensities from Affymetrix SNP6 .CEL files for tumor samples across different cancer types were combined, calibrated, normalized, and segmented in a uniform fashion as previously described[50]. Markers identified as having recurrent germline copy number variations using normal samples were excluded. For samples for which ABSOLUTE[54] purity and ploidy calls were available, copy numbers were scaled by a factor inversely proportional to the purity estimate to remove the effects of admixed normal cells, as previously described[50]. Copy number profiles were deconstructed into underlying somatic events using GISTIC[55]; only focal events with length less than 1 Mb were used, to limit normalization bias between males and females on the X chromosome. Copy number for a gene was determined using the most extreme copy number among the markers spanning the gene. The threshold for calling somatic copy number events was chosen by considering the distribution of copy number across both male and female cohorts (**Supplementary Fig. 3**). Because normalization decouples overall copy number for the X chromosome from that for the autosomes, retention or loss of the entire X chromosome was called by applying a threshold of 1.6 to a robust average of the non-normalized, calibrated signal across the X chromosome (**Supplementary Fig. 3**). The maximum and minimum 0.2% of probe-level signals were excluded.

Power analysis. Power calculation was performed using a binomial model similarly as previously described[39]. The power to detect a male-biased mutation can be determined by first calculating the maximum number of mutations in males, m_{max}, that would be considered non-significant in the null model, based on an inverse binomial cumulative distribution function, given the desired bound on the P value (P < 0.1/number of genes), the total number of mutations in a gene in the entire cohort, n (a function of the fraction of patients with a mutation and the total number of patients), and the fraction of all events that occur in males (as opposed to females) in the cohort, which determines the probability that a mutation will be in a tumor in a male, p_0. The power is then determined by the probability of discovering at least that many mutations in male patients under the alternate hypothesis, which adjusts p_0 by the hypothesized increased relative

risk of a mutation occurring in a tumor in a male, R_{signal}, the fold change in comparison to the null hypothesis.

$$R_{background} = p_0/(1-p_0)$$

$$R_{total} = R_{background} \times R_{signal}$$

$$p_{alt} = 1/(1+1/R_{total})$$

To generate figures, we calculated the number of patients (N) required for power to detect 80% of genes with a statistically significant male bias, given hypothetical values of R_{signal}, n, and p_0 ($R_{signal} = 4$ in **Fig. 3a**).

Y-chromosome copy loss determination. We calculated Y-chromosome coverage in megabase bins for paired tumor and normal sequencing data and then normalized to total coverage of the exome. A region of the Y chromosome at bases 29,000,001–58,000,000 was omitted because of poor coverage. An additional region (13,000,001–14,000,000) was removed because of high incidence of misalignment. A biphasic pattern of copy number quantification was observed in male samples (**Supplementary Fig. 6**), with a distinct population of tumors demonstrating significant median copy loss in comparison to paired normal samples. Y-chromosome loss was therefore assigned to those samples with <25% the overall coverage of the Y chromosome in the tumor as compared to the paired normal sample.

Allele-specific expression analysis. Exome sequencing and RNA–seq data from somatic and germline samples of GBM, LGG, HNSC, KIRC, LUSC, and LUAD sets from TCGA and brain, lung, and whole-blood sample sets from GTEx were analyzed. For each, allele-specific RNA–seq pileup counts were called at heterozygous germline sites and tumor sites from cancers (determined by exome sequencing) for the respective analysis. Duplicate, non-primary, soft-clipped and Phred quality 0 reads were not included in the pileup count. For sites where at least 20 reads (tumor samples) or 8 reads (normal samples) were detected by RNA–seq, the count of the less frequent allele was divided by the total allele count to obtain the minor allele fraction. Data are represented as the average minor allele fraction for all sites in the indicated gene or as the average of all sites in all genes in the non-escape group.

Analysis of GTEx male versus female expression data. We obtained RNA–seq expression data from the GTEx Project as gene reads per sample by accessing the project's pipeline (http://www.gtexportal.org/home/). For each tissue, we normalized the signal across samples. The whole-blood samples used were restricted to those collected ante mortem. There were 85 whole-blood samples (25 female and 60 male) and 360 brain samples (124 female and 233 male) that passed quality control and fit the selection criteria. We tested for bimodality of *ATRX* expression in the brain in two ways. First, we fit a Gaussian mixture model to the empirical densities using the R package mixtools (http://cran.r-project.org/web/packages/mixtools/), function normalmixEM with parameters k = 2, epsilon = 1e-08, maxit = 1,000, maxrestarts = 20, and assessed the distance between resulting means. Second, we tested for bimodality with a likelihood-ratio test for bimodality in two-component mixtures. The method contrasts the likelihood of the data obtained under restricted and unrestricted maximum-likelihood fits of mixture of normal distributions. Under the assumption of equal variance, bimodality solely depends on mixture component weight and the ratio of the distance between the means and the variance. With unequal variance, it is also determined by the ratio of variances. We used R package diptest to test for multimodality (http://cran.r-project.org/web/packages/diptest/).

Calculation of the percentage excess male risk. For a given gene in a specific disease, we calculated the excess male risk associated with loss-of-function mutation of a single gene on the X chromosome. We cannot assume that the M:F ratio in our sample set is the same as the general population incidence of that disease, so we adjusted the overall M:F ratio to SEER data for each cancer in the US population. The fraction of the excess male risk in the disease attributable to a specific gene mutation was calculated as follows: (n males with the gene mutated in our data set – ($Z \times n$ females with the gene mutated in our data set))/(n males in our data set – ($Z \times n$ females in our data set)), where Z is our data set M:F ratio divided by the SEER data M:F ratio.

***Cnksr2* knockdown, immunoblotting, and soft agar colony assays.** Mouse *Cnksr2* or control *RFP* shRNA constructs (oligonucleotide sequences in **Supplementary Table 12**) were cloned according to the protocol at http://www.addgene.org/tools/protocols/plko/ into a pLKO.1 vector modified to express a GFP reporter in place of the puromycin resistance gene. Lentivirus was produced, and mouse NIH 3T3 cells (from the American Type Culture Collection, mycoplasma free) were infected using standard protocols (Addgene). GFP-positive cells were obtained by sorting after 7 d. Immunoblotting was performed as previously described[56] using antibodies recognizing phosphorylated ERK (Cell Signaling Technology, 4370), total ERK (Cell Signaling Technology, 9102), and tubulin (Sigma, T6074). All primary antibodies were used at 1:1,000 dilutions. Soft agar colony assays were performed as previously described[57]. After 3 weeks, random microscopy fields with an area of 3.29 mm^2 were scanned and colonies of at least 50 μm^2 in size were counted using a CellCelector (ALS, Germany).

RNA sequencing and gene set enrichment analysis in *Cnksr2* knockdown cells. Total RNA was prepared using a miRNeasy kit (Qiagen). Illumina sequencing libraries were prepared using Illumina TruSeq Stranded mRNA sample preparation kits from 500 ng of purified total RNA according to the manufacturer's protocol. The finished double-stranded DNA libraries were quantified by Qubit fluorometer, Agilent TapeStation 2200, and RT–qPCR using the Kapa Biosystems library quantification kit according to the manufacturer's protocols. Uniquely indexed libraries were pooled in equimolar ratios and sequenced on an Illumina NextSeq 500 with single-end 75-bp reads by the Dana-Farber Cancer Institute Molecular Biology Core Facilities. Reads were aligned to the mm9 reference genome assembly using STAR (v25.1b) (https://github.com/alexdobin/STAR). FPKM expression values were calculated using Cufflinks (v2.2.1) (http://cole-trapnell-lab.github.io/cufflinks/). Spearman correlation and principal-component analysis were performed using VIPER (https://bitbucket.org/cfce/viper/). GSEA (http://www.broadinstitute.org/gsea/) was performed as previously described[41,58]. Network enrichment mapping was performed using Cytoscape (http://www.cytoscape.org/).

50. Zack, T.I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140 (2013).
51. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
52. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
53. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
54. Carter, S.L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
55. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl. Acad. Sci. USA* **104**, 20007–20012 (2007).
56. Yoda, A. *et al.* Mutations in G protein β subunits promote transformation and kinase inhibitor resistance. *Nat. Med.* **21**, 71–75 (2015).
57. Hammerman, P.S. *et al.* Mutations in the *DDR2* kinase gene identify a novel therapeutic target in squamous cell lung cancer. *Cancer Discov.* **1**, 78–89 (2011).
58. Lane, A.A. *et al.* Triplication of a 21q22 region contributes to B cell transformation through *HMGN1* overexpression and loss of histone H3 Lys27 trimethylation. *Nat. Genet.* **46**, 618–623 (2014).

doi:10.1038/ng.3726