Sex-Associated Gene Expression Alterations Correlate With Esophageal Cancer Survival

Nathaniel Weygant, PhD[1], Karen Chang, DO[2], Christian S. Jackson, MD[3] and Kenneth J. Vega, MD[4]

OBJECTIVES: Esophageal cancer (EC) is a significant cause of cancer death with 5-year survival of 10%–15% and males more frequently affected. Genetic evaluation for loci highlighting risk has been performed, but survival data are limited. The Cancer Genome Atlas (TCGA) data sets allow for potential prognostic marker assessment in large patient cohorts. The study aimed to use the TCGA EC data set to assess whether survival varies by sex and explore genetic alterations that may explain variation observed.

METHODS: TCGA clinical/RNA-seq data sets (n = 185, 158 males/27 females) were downloaded from the cancer genome browser. Data analysis/figure preparation was performed in R and GraphPad Prism 7. Survival analysis was performed using the survival package. Text mining of PubMed was performed using the tm, RISmed, and wordcloud packages. Pathway analysis was performed using the Reactome database.

RESULTS: In EC, male sex/high tumor grade reduced overall survival (hazard ratio = 2.27 [0.99–5.24] for M vs F and 2.49 [0.89–6.92] for low vs high grade, respectively) and recurrence-free survival (hazard ratio = 4.09 [0.98–17.03] for M vs F and 3.36 [0.81–14.01] for low vs high grade, respectively). To investigate the genetic basis for sex-based survival differences in EC, corresponding gene expression data were analyzed. Sixty-nine genes were dysregulated at the $P < 0.01$ level by the Wilcox test, 33% were X-chromosome genes, and 7% were Y-chromosome genes.

DISCUSSION: Female sex potentially confers an EC survival advantage. Importantly, we demonstrate a genetic/epigenetic basis for these survival differences that are independent of lifestyle-associated risk factors overrepresented in males. Further research may lead to novel concepts in treating/measuring EC aggressiveness by sex.

SUPPLEMENTARY MATERIAL accompanies this paper at http://links.lww.com/CTG/A472 and http://links.lww.com/CTG/A473.

INTRODUCTION

Esophageal cancer (EC) in 2020 is estimated to be the sixth most common digestive cancer diagnosed and fourth most frequent cause of digestive cancer death in the United States (1). Besides, 5-year survival for this cancer remains dismal at 20% (1). EC has 2 distinct subtypes, squamous cell carcinoma and adenocarcinoma, each with unique incidence patterns. Risk factors for squamous cell carcinoma include smoking and alcohol intake while gastroesophageal reflux disease and obesity increase adenocarcinoma risk (2,3). Both cancers have a male predominance, more significant for adenocarcinoma than squamous cell carcinoma (4). Males are also more likely to die from EC or have decreased survival once diagnosed compared with females (5–8). Moreover, males have inferior outcomes in EC with or without surgical

intervention (9,10). Genetic variants for EC cancer risk are modest at best (11). Limited data exist assessing the impact of genetic variations present between the sexes on survival in EC. This investigation aimed to evaluate EC survival by sex and explore potential genetic alterations that may explain any difference seen using a known cancer database.

METHODS

The Cancer Genome Atlas data sets

Clinical and mRNA expression data sets derived from EC generated by The Cancer Genome Atlas (TCGA) Consortium were downloaded from the UCSC Xena project (xenabrowser.net). A total of 185 patients including both squamous cell and adenocarcinoma histological types of EC were assessed. The data set

ESOPHAGUS

Table 1. Clinical patient characteristics for the TCGA ESCA dataset

	Male	Female	P
Sample size (n)	158	27	
Age (median)	60	67	
Race, n (%)			0.232
Not reported	17 (11)	3 (11)	
White	93 (59)	21 (78)	
Asian	43 (27)	3 (11)	
Black or African American	5 (3)	0 (0)	
History of reflux, n (%)			0.172
Unknown	28 (18)	3 (11)	
No	85 (54)	12 (44)	
Yes	45 (28)	12 (44)	
Barrett's esophagus, n (%)			1.000
Unknown	35 (22)	4 (15)	
No	99 (63)	19 (70)	
Yes	24 (15)	4 (15)	
Tobacco history, n (%)			0.275
Unknown	18 (11)	1 (4)	
Lifelong nonsmoker	44 (28)	12 (44)	
Current reformed smoker (<15 yr)	33 (21)	4 (15)	
Current reformed smoker (>15 yr)	30 (19)	6 (22)	
Current smoker	33 (21)	4 (15)	
Alcohol history, n (%)			0.001
Unknown	61 (39)	11 (41)	
0 drinks per day	22 (14)	12 (44)	
1 drink per day	10 (6)	1 (4)	
2 or more drinks per day	65 (41)	3 (11)	
Histological type, n (%)			0.835
Adenocarcinoma, NOS	77 (49)	12 (44)	
Squamous cell carcinoma	81 (51)	15 (56)	
Tumor stage, n (%)			0.022
Stage I	12 (8)	7 (28)	
Stage II	66 (48)	13 (52)	
Stage III	52 (38)	4 (16)	
Stage IV	8 (6)	1 (4)	
Tumor grade, n (%)			0.135
GX	32 (20)	8 (30)	
G1	14 (9)	5 (19)	
G2>	112 (71)	14 (52)	
Distant metastasis, n (%)			1.000
MX	35 (22)	5 (19)	
M0	115 (73)	21 (78)	
M1	8 (5)	1 (4)	

Table 1. (*continued*)

	Male	Female	P
Adjuvant chemotherapy, n (%)			0.418
Unknown	13 (8)	3 (11)	
No	112 (71)	21 (78)	
Yes	33 (21)	3 (11)	
Surgery status, n (%)			0.438
Unknown	92 (58)	21 (78)	
Yes	52 (33)	5 (19)	
No	14 (9)	1 (3)	

ESCA, Esophageal Carcinoma; NOS, NOS, not otherwise specified; TCGA, The Cancer Genome Atlas.
When comparing clinical factors by patient sex, the TCGA dataset demonstrates relatively similar patient histories and tumor characteristics except for alcohol history (P = 0.001 by Fisher's exact test).

included 27 female and 158 male patients. Through the study period, 8 female and 18 male patients were lost to follow-up, and 2 male patients were reported to have received treatment (chemotherapy/radiotherapy combination) before surgery. Patient samples were procured in the United States (n = 77), Vietnam (n = 42), Brazil (n = 17), Canada (n = 14), the Netherlands (n = 12), Russia (n = 12), Ukraine (n = 6), Australia (n = 1), the United Kingdom (n = 1), Bulgaria (n = 1), and unreported (n = 2). Races of patients were white, Asian, and black in order of frequency in the data set. The average and median length of follow-up were 19 and 13 months, respectively. Further information about the data set can be obtained from TCGA.

Univariate and multivariate survival analyses
The Cox proportional hazards model and Kaplan-Meier analyses were used to assess differences in overall (OS) and recurrence-free survival (RFS) between selected groups. OS was considered the primary outcome for the study. Death during the study period from any cause was classified as an OS event. Patients who were lost to follow-up or living patients were censored at the time point of their last follow-up according to the information collected in the data set. Univariate analyses were performed in R using the coxph and survdiff functions from the "survival" package. For multivariate survival analyses, the coxph function was used to assess the independence of the clinical factors determined to be significant in univariate analyses.

Heatmaps
Heatmaps were prepared in R using the heatmap.2 function from the "gplots" package. Genes dysregulated between tumors of male and female patients where determined using the Wilcox test (*wilcox.test* function). The hierarchical clustering of patients has performed automatically within sex categories while the clustering of gene expression was performed across sex.

Word cloud meta-analysis
To obtain meta-data for word cloud analysis, the RISmed package was used to scrape PubMed for abstracts for all articles mentioning a

Table 2. Univariate and multivariate analysis of survival based on clinical patient characteristics from the TCGA ESCA dataset

Clinical factor	Univariate analysis				Multivariate analysis			
	HR (95% CI) OS	P (OS)	HR (95% CI) RFS	P (RFS)	HR (95% CI) OS	P (OS)	HR (95% CI) RFS	P (RFS)
Gender (male vs female)	**2.27 (0.99–5.24)**	**0.031**	**4.09 (0.98–17.03)**	**0.017**	4.50 (1.02–19.93)	0.048	6.35 (0.85–47.37)	0.072[a]
Age	1.00 (0.98–1.02)	0.876	1.00 (0.97–1.02)	0.747				
History of reflux (yes vs no)	1.05 (0.62–1.76)	0.866	0.92 (0.44–1.89)	0.810				
Barrett's esophagus (yes vs no)	1.04 (0.59–1.82)	0.899	1.20 (0.45–3.22)	0.724				
Tobacco history								
Smoker vs nonsmoker	1.44 (0.82–2.53)	0.193	1.00 (0.51–1.96)	0.996				
Reformed smoker vs nonsmoker	1.32 (0.72–2.41)	0.366	0.89 (0.42–1.89)	0.758				
Current smoker vs nonsmoker	1.76 (0.87–3.57)	0.114	1.25 (0.52–2.98)	0.615				
Alcohol history								
Drinker vs nondrinker	2.30 (0.81–6.53)	0.084	1.11 (0.47–2.63)	0.811				
1 vs 0 drinks per day (avg)	2.08 (0.68–6.33)	0.196	0.72 (0.26–2.00)	0.534				
>1 vs 0 drinks per day (avg)	2.53 (0.84–7.60)	0.098	1.72 (0.68–4.37)	0.256				
Histological type (EAC vs ESCC)	0.88 (0.55–1.40)	0.581	1.48 (0.75–2.94)	0.256				
Mucosal dysplasia (present vs absent)	1.31 (0.71–2.44)	0.388	2.98 (0.92–9.71)	0.057				
Intestinal metaplasia (present vs absent)	1.39 (0.49–3.98)	0.552	9.22 (2.64–32.14)	0.001				
Tumor grade								
High vs low	2.49 (0.89–6.92)	0.047	3.36 (0.81–14.05)	0.046	0.83 (0.27–2.56)	0.745	2.41 (0.55–10.67)	0.246
G2 vs G1	2.14 (0.75–6.14)	0.157	3.04 (0.70–13.10)	0.136	1.71 (0.59–4.99)	0.32	2.22 (0.49–10.11)	0.301
G3 vs G1	2.98 (1.03–8.56)	0.043	3.91 (0.89–17.22)	0.072	2.45 (0.84–7.11)	0.099	2.87 (0.59–13.91)	0.190
Tumor stage								
Adv. vs early	3.20 (1.90–5.38)	<0.001	0.89 (0.40–2.01)	0.800	2.49 (0.89–6.92)	0.047	0.69 (0.29–1.61)	0.388
Stage II vs stage I	1.85 (0.70–4.92)	0.216	1.71 (0.51–5.81)	0.387	3.63 (0.74–17.86)	0.113	1.90 (0.24–14.92)	0.541
Stage III vs stage I	4.70 (1.72–12.82)	0.003	1.49 (0.39–5.62)	0.559	9.06 (1.72–47.73)	0.009	1.27 (0.15–11.10)	0.828
Stage IV vs stage I	9.99 (3.09–32.31)	<0.001	N/A	N/A	14.51 (2.26–93.01)	0.005	N/A	N/A
Adjuvant chemotherapy (treated vs untreated)	1.10 (0.59–2.07)	0.763	1.11 (0.50–2.50)	0.56				

CI, confidence interval; EAC, esophageal adenocarcinoma; ESCA, Esophageal Carcinoma; ESCC, esophageal squamous cell carcinoma; HR, hazard ratio; OS, overall survival; RFS, recurrence-free survival; TCGA, The Cancer Genome Atlas.
Univariate cox regression analysis demonstrated that sex and tumor grade predicted both overall and recurrence-free survival in esophageal cancer (P < 0.05). Additionally, intestinal metaplasia as determined by the presence of goblet cells significantly predicted recurrence in esophageal cancer (P = 0.001). [a]Multivariate cox regression analysis demonstrates the marginally insignificant prognostic potential for sex in esophageal cancer (P < 0.057).

primary search term and "cancer." The resulting abstracts were manipulated and cleaned of common words and baseline words (4,000 most common words contained in abstracts mentioning the search term cancer) using the tm text-mining package. The resulting most common, relevant terms were plotted using the wordcloud package in R. Finally, the top 10 genes determined by text mining were subjected to further pathway enrichment analysis in Reactome to discover additional pathways of interest.

Data analysis

Clinical and demographic categories for the Fisher exact test and survival analyses were selected *a priori*. All statistical tests were performed in R v3.2 and GraphPad Prism 7. For nonparametric data, the Wilcox test was used to make comparisons. Pathway analyses were performed using the Reactome database (12). All findings $P < 0.05$ were considered significant.
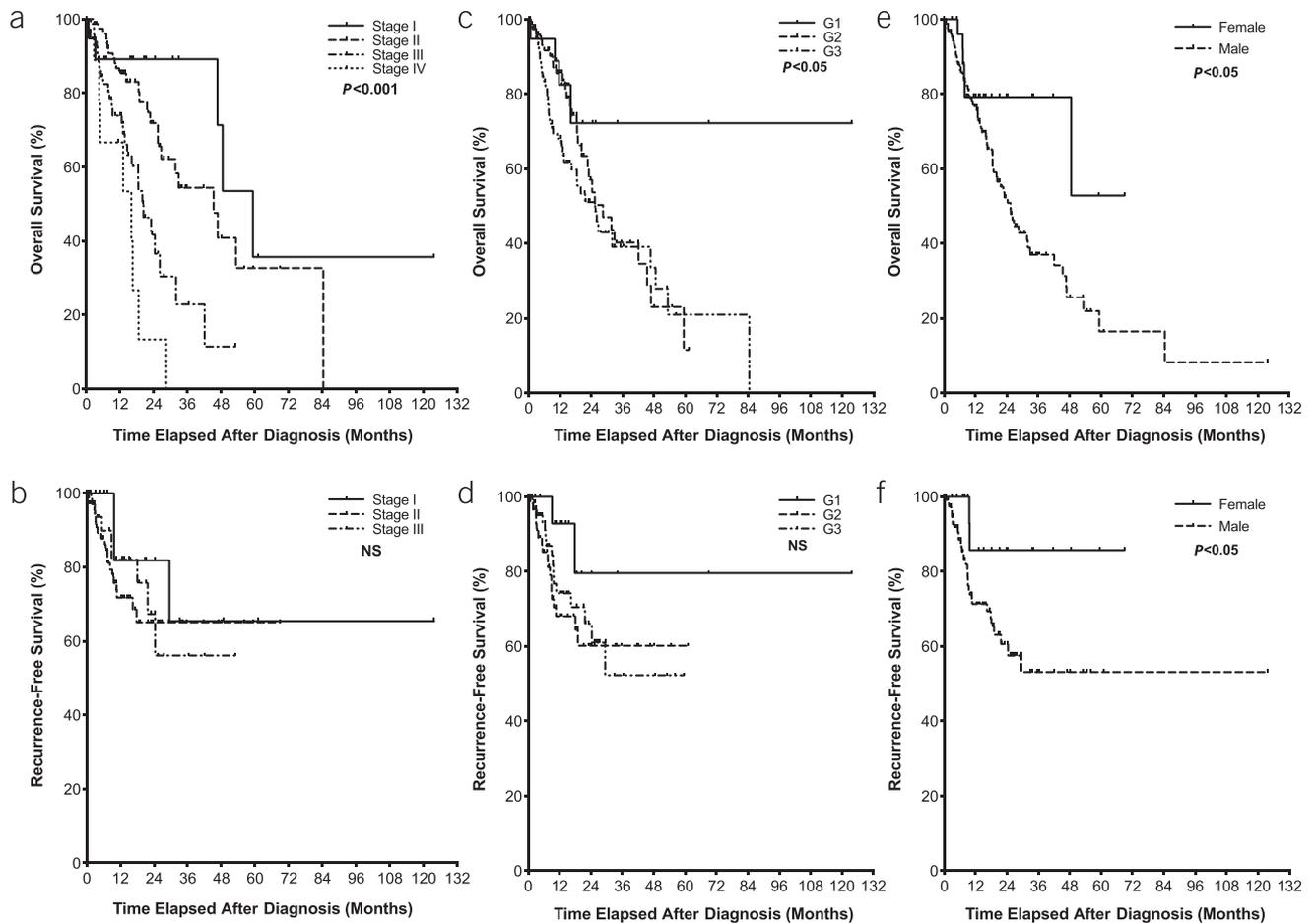
ESOPHAGUS

Figure 1. Kaplan-Meier survival analysis demonstrates prognostic significance for tumor stage, tumor grade, and sex in esophageal cancer survival. (**a** and **b**) Tumor stage predicts overall survival but not recurrence-free survival in esophageal cancer. (**c** and **d**) Tumor grade predicts overall survival but not recurrence-free survival in esophageal cancer. (**e** and **f**) Sex predicts both overall survival and recurrence-free survival in esophageal cancer.

RESULTS

Clinical patient characteristics from TCGA ESCA

TCGA Esophageal Carcinoma (ESCA) data set contains a total of 185 patients. Among male patients (n = 158), the median age was 60 years compared with 67 years for female patients (n = 27). The prevalence or absence of relevant clinical factors including tobacco and alcohol consumption history, history of gastroesophageal reflux disease, Barrett's esophagus, histological type (squamous cell or adenocarcinoma), tumor grade, tumor stage, lymph node and distant metastasis status, adjuvant therapy, and surgical therapy were compared between male and female patients using the Fisher exact test (Table 1). All comparisons were insignificant except alcohol consumption and tumor stage. Alcohol consumption was higher on average among male patients ($P = 0.001$).

Sex is a potential independent prognostic factor in EC

To understand the clinical basis for survival differences in the TCGA EC patient population, Cox regression analyses were performed using available clinical and histologic factors (Table 2). Among the selected factors, male sex, tumor grade, and tumor stage led to significantly reduced overall survival with hazard ratios of 2.27 (0.99–5.24; M vs F), 2.49 (0.89–6.92; high vs low grade), and 3.20 (1.90–5.38; advanced vs early stage), respectively.

Sex and grade also predicted recurrence-free survival with a hazard ratio of 4.09 (0.98–17.03) for males compared with females and 3.36 (0.81–14.05) for high-grade compared with low-grade tumors. In addition, the appearance of intestinal metaplasia, as indicated by the presence of goblet cells, was highly predictive of recurrent disease (hazard ratio [HR]: 9.22 [2.64–32.14]; intestinal metaplasia was present vs absent; $P = 0.001$) in EC. Since intestinal metaplasia is a factor almost exclusively associated with adenocarcinoma, this analysis was repeated in the adenocarcinoma patient subset alone revealing similar findings (HR: 7.29 [1.66–32.05]; intestinal metaplasia present vs absent; $P = 0.009$). To better understand our findings, we performed Kaplan-Meier survival analysis for sex and tumor grade. The log-rank test confirmed our findings concerning differential overall and recurrence-free survival by sex and overall survival by tumor grade and stage (Figure 1).

Considering the findings of other groups demonstrating that sex is an independent prognostic factor in EC (13), we performed multivariate Cox regression analysis. Because the number of patients with intestinal metaplasia was limited and the sample size was low for the female sex, we were unable to include this factor in the analysis. The Cox proportional hazards model for the combination of sex, tumor grade, and tumor stage highlighted the prognostic significance for sex
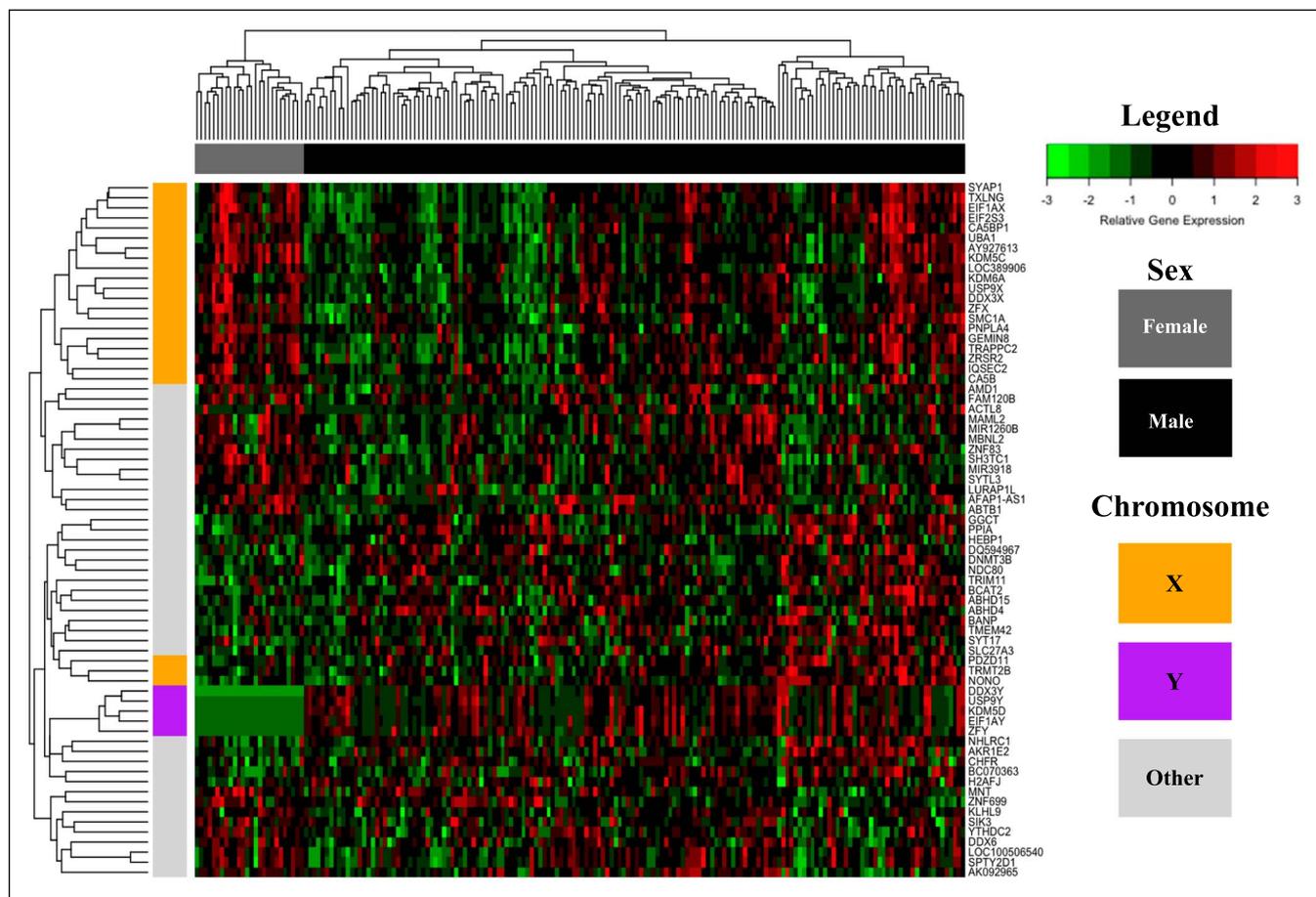
Figure 2. Select genes are differentially expressed between female and male patient tumors. Analysis of RNA-seq data from The Cancer Genome Atlas esophageal carcinoma data set demonstrates a differential expression of 69 genes at the $P < 0.01$ significance level (Wilcox test). Among these genes are a significant number of X-chromosome genes most of which are overexpressed in female patient tumors and a handful of Y-chromosome genes overexpressed in male patient tumors (black columns).

and tumor stage in predicting OS in the TCGA ESCA data set. With inclusion of stage IV tumors, it was not possible to fit either univariate or multivariate Cox proportional hazards models for analysis of RFS. However, multivariate analysis including stages I-III revealed that sex was the most strongly associated factor with RFS (HR: 6.35 [0.85–47.37]; M vs F) (Table 2).

Sex-specific gene alterations in esophageal tumors are enriched from X and Y chromosomes

To investigate whether there is a genetic basis for survival differences by sex in EC, we set out to analyze the corresponding TCGA RNA-Seq expression data. In total, 326 genes were differentially expressed when comparing male with female patient tumors ($P < 0.05$; see Supplementary Figure 1, Supplementary Digital Content 1, http://links.lww.com/CTG/A472). To narrow down candidate genes, we further investigated genes at the $P < 0.01$ significance level (n = 69). Notably, 33% of these genes were expressed from the X chromosome (n = 23), and 7% were Y-chromosome genes (n = 5) expressed in male patient tumors (Figure 2). Eighty-seven percent of dysregulated X-chromosome genes in our analysis were overexpressed in female compared with male patient tumors. Interestingly, 35% of these genes, all of

which were overexpressed in female patient tumors in our data set, were previously reported to escape X-inactivation[1]—the process by which the additional X chromosome in females is silenced to prevent undue influence on gene expression. Moreover, 5 of these genes (DDX3X, EIF1AX, KDM5C, USP9X, and ZFX) are the X-chromosome homologs of the Y-chromosome genes significantly expressed in male tumors (DDX3Y, EIF1AY, KDM5D, USP9Y, and ZFY). To investigate the expression of the select sex-associated genes in EC histological subtypes, we reproduced gene expression heatmaps in the esophageal adenocarcinoma and esophageal squamous cell carcinoma subtypes separately. The pattern of differential expression in both subtypes was comparable with the total EC data set (see Supplementary Figure 2, Supplementary Digital Content 2, http://links.lww.com/CTG/A473).

Sex-specific gene alterations in esophageal tumors favor nontranscriptional pathways

To assess the ontological significance of the altered genes, we performed pathway enrichment analysis based on gene identity using the Reactome database. In total, 19 pathways were significantly ($P < 0.05$) enriched from the gender-specific gene expression data, including 8 epigenetic pathways, 6 proteasomal

ESOPHAGUS

Table 3. Pathway analysis of sex-dysregulated genes in esophageal cancer

Reactome ID	Pathway name	Fraction	P	FDR
R-HSA-3214858	RMTs methylate histone arginines	11/49	2.32E-13	5.06E-11
R-HSA-3247509	Chromatin modifying enzymes	14/241	5.60E-09	3.26E-07
R-HSA-4839726	Chromatin organization	14/241	5.60E-09	3.26E-07
R-HSA-5689901	Metalloprotease DUBs	7/31	6.04E-09	3.26E-07
R-HSA-3214815	HDACs deacetylate histones	6/60	8.31E-06	3.57E-04
R-HSA-5689603	UCH proteinases	6/96	1.12E-04	0.00401877
R-HSA-74217	Purine salvage	3/13	1.55E-04	0.00481002
R-HSA-3214847	HATs acetylate histones	6/108	2.10E-04	0.00567509
R-HSA-5689880	Ub-specific processing proteases	7/205	0.00113161	0.02699193
R-HSA-3214842	HDMs demethylate histones	3/27	0.00128533	0.02699193
R-HSA-5688426	Deubiquitination	8/280	0.00155834	0.02960854
R-HSA-8866652	Synthesis of active ubiquitin: roles of E1 and E2 enzymes	3/30	0.00173386	0.0312095
R-HSA-73847	Purine metabolism	3/34	0.00246832	0.03949309
R-HSA-1475029	Reversible hydration of carbon dioxide	2/12	0.00406326	0.06094883
R-HSA-8852135	Protein ubiquitination	3/58	0.01072502	0.15015034
R-HSA-5334118	DNA methylation	2/34	0.02922007	0.36433493
R-HSA-6803544	Ion influx/efflux at host-pathogen interface	1/4	0.03068715	0.36433493
R-HSA-15869	Metabolism of nucleotides	3/99	0.04258907	0.36433493
R-HSA-212300	PRC2 methylates histones and DNA	2/42	0.04285481	0.36433493

Pathway analysis of genes dysregulated at the $P < 0.01$ level demonstrates specific enrichment of interrelated epigenetic/methylation and proteasomal/ubiquitination-related pathways. In addition, purine-processing pathways were enriched. FDR, false discovery rate.

degradation pathways, 3 purine/nucleotide metabolism pathways, and 2 additional pathways for carbonic anhydrases and host/pathogen interactions (Table 3). To expand the analysis to capture other key pathways, we mined PubMed for all abstracts related to the 12 pathways with adjusted $P < 0.05$ by false discovery rate. After normalizing the abstracts by removing English stop words and generic terms as described in the methods, we performed pathway enrichment analysis with the terms receiving the most mentions in the combined abstracts. In addition to significant enrichment for essential epigenetic pathways noted in the previous analysis, the most significant pathway enriched was the NOTCH signaling pathway (Table 4), which has been found to drive tumorigenesis in esophageal and other gastrointestinal cancers, though controversies remain (14).

DISCUSSION

The present investigation on EC was performed using the TGCA data set to determine whether survival varies based on sex and explore any potential genetic alterations related to such variation if found. The results indicate that female sex may be favorable as well as an independent prognostic factor potentially conferring improved overall and disease-free survival in EC compared with males due to X-chromosome inactivation (XCI). Sex seems to play a role in the survival of EC patients through genetic factors rather than lifestyle differences present between the sexes alone. Determination of the relative contribution of these genetic factors to survival in

comparison with lifestyle factors will need to be assessed in larger studies with additional preselected criteria such as body mass index. Earlier studies have suggested that gender may be prognostic only for squamous cell histology or for certain stages of squamous cell carcinomas (13,15,16). However, there was an equal distribution of adenocarcinoma and squamous cell carcinoma in the TCGA database, allowing us to match multiple clinical factors between both male and female populations.

XCI has been known to occur unequally in a population of somatic cells (17). It was previously hypothesized that one X chromosome in female somatic cells (Xi) is randomly inactivated by epigenetic factors during the embryonic stage (18). The activated X chromosome (Xa) carries on its transcription as normal, while the Xi is largely silenced, protecting the cell against possible overexpression of oncogenes (19). Further research demonstrated that not only is the Xi frequently incompletely inactivated but also a portion of these X-chromosome escape genes transcribe oncogenes or other factors promoting carcinogenesis (20). It has also been reported that the Xi-chromosome genes that escape inactivation have been noted to have Y homologs, which cause the Xi chromosome to act similarly to male XY chromosome (21). The active Xa can contain genes that mitigate the tumor-promoting effect in the female somatic cell, but the Y chromosome in male somatic cells may not have the same failsafe system (22). In our study, we found that in female tumors, 87%

Table 4. Additional enriched pathways associated with sex in esophageal cancer

Reactome ID	Pathway name	Fraction	*P*	FDR
R-HSA-2122947	NOTCH1 intracellular domain regulates transcription	5/48	3.59E-09	5.05E-07
R-HSA-2894858	Signaling by NOTCH1 HD + PEST domain mutants in cancer	5/58	9.19E-09	5.05E-07
R-HSA-2644602	Signaling by NOTCH1 PEST domain mutants in cancer	5/58	9.19E-09	5.05E-07
R-HSA-2894862	Constitutive signaling by NOTCH1 HD + PEST domain mutants	5/58	9.19E-09	5.05E-07
R-HSA-2644606	Constitutive signaling by NOTCH1 PEST domain mutants	5/58	9.19E-09	5.05E-07
R-HSA-2644603	Signaling by NOTCH1 in cancer	5/58	9.19E-09	5.05E-07
R-HSA-1980143	Signaling by NOTCH1	5/73	2.87E-08	1.35E-06
R-HSA-1368082	RORA activates gene expression	4/27	4.05E-08	1.66E-06
R-HSA-193670	p75NTR negatively regulates cell cycle via SC1	3/6	6.47E-08	2.39E-06
R-HSA-157118	Signaling by NOTCH	5/111	2.27E-07	7.50E-06
R-HSA-212165	Epigenetic regulation of gene expression	5/117	2.94E-07	7.91E-06
R-HSA-1989781	PPARA activates gene expression	5/120	3.33E-07	7.91E-06
R-HSA-4839726	Chromatin organization	6/241	3.44E-07	7.91E-06
R-HSA-3247509	Chromatin modifying enzymes	6/241	3.44E-07	7.91E-06
R-HSA-400206	Regulation of lipid metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha)	5/123	3.76E-07	8.28E-06
R-HSA-3214815	HDACs deacetylate histones	4/60	9.62E-07	1.92E-05
R-HSA-400253	Circadian clock	4/68	1.58E-06	3.00E-05
R-HSA-5250913	Positive epigenetic regulation of rRNA expression	4/75	2.32E-06	4.18E-05
R-HSA-381340	Transcriptional regulation of white adipocyte differentiation	4/84	3.63E-06	6.17E-05
R-HSA-5619507	Activation of HOX genes during differentiation	4/91	4.98E-06	7.46E-05

Using the final word cloud data, genes were isolated and subjected to pathway enrichment analysis in Reactome. Pathways of note include NOTCH signaling pathways and epigenetic pathways. FDR, false discovery rate.

ESOPHAGUS

of dysregulated genes were overexpressed compared with these same genes in the male tumors. Over a third of these dysregulated genes have been known to escape XCI, and several dysregulated genes also have a Y-chromosome homolog, subjecting males 2 "hits" in the Knudson cancer hypothesis (19).

Of the genes we found that were significantly more likely to be dysregulated in tumor cells, 5 of them are known to escape XCI (23). KDM5C is a histone H3 lysine demethylase, DDX3X has both nuclear and cytoplasmic roles, and its dysregulation has been implicated in tumorigenesis. EIF1AX is an essential eukaryotic translation initiation factor, and its mutated form has been associated with papillary and anaplastic thyroid carcinoma. USP9X encodes a protein that is similar to ubiquitin-specific proteases that regulate the degradation of proteins (24–27). ZFX plays a critical role in the maintenance of self-renewal in embryonic stem cells, and mutations have been linked to several cancers including colorectal, hepatocellular, and oral squamous

carcinoma (28,29). Of special interest, the NOTCH signaling pathway maintains stem cells in the gastrointestinal tract and its dysregulation is implicated in the progression of both esophageal adenocarcinoma and squamous cell carcinoma (30). Taken together, there is a strong suggestion that the significant difference seen between males and females in incidence and survival of EC may be attributed to X-linked genes, particularly those that escape normal XCI.

This study has several limitations. Despite the size of TCGA database, only a limited number of patients with a diagnosis of EC were available for analysis. The small sample size, inclusion of multiple histological types, and other factors could lead to false-positive associations, and it will be essential for further studies to be conducted in larger data sets. Therefore, we urge caution about drawing strong conclusions from this analysis, but also note that, given the relative dearth of genetic data for EC, it is important to make use of existing resources to guide future studies. In addition, only 15% of the patients were female, which potentially could

limit the power of the current investigation. However, this is consistent with the epidemiology of EC (4). Moreover, although a minority of patients received adjuvant chemotherapy in the study population, actual regimens used were not available for review. This may have impacted survival due to patient response or lack thereof. Finally, the statistically significant association of intestinal metaplasia with disease-free survival in adenocarcinoma is interesting and should be investigated in larger studies, as this is easily assessed by pathology and could be a useful indicator of recurrence risk.

In conclusion, female sex seems to be a favorable independent prognostic factor, conferring improved overall and disease-free survival in EC compared with males, potentially due to XCI. Understanding how XCI variation impacts treatment decisions and outcomes are essential to provide the optimal result for EC patients.

CONFLICTS OF INTEREST

Guarantor of the article: Kenneth J. Vega, MD.
Specific author contributions: N.W. and K.J.V.: planned the study. N.W., K.C., and K.J.V.: wrote the manuscript. All authors interpreted the data, revised the manuscript for intellectual content, and approved the final manuscript.
Financial support: None to report.
Potential competing interests: None to report.

Study Highlights

WHAT IS KNOWN

✓ Esophageal carcinomas (squamous cell and adenocarcinoma) primarily affect males.
✓ Survival from esophageal carcinomas is poor.

WHAT IS NEW HERE

✓ Sex-specific genetic alterations may confer survival advantage for esophageal carcinomas independent of modifiable lifestyle factors.

TRANSLATIONAL IMPACT

✓ Genetic and epigenetic factors may play a more significant role in survival than previously known.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7–30.
2. Rubenstein JH, Shaheen NJ. Epidemiology, diagnosis and management of esophageal adenocarcinoma. Gastroenterology 2015;149:302–17.
3. Abnet CC, Arnold M, Wei WQ. Epidemiology of esophageal squamous cell carcinoma. Gastroenterology 2018;154:360–73.
4. Islami F, DeSantis CE, Jemal A. Incidence trends of esophageal and gastric cancer subtypes by race, ethnicity and age in the United States, 1997–2014. Clin Gastroenterol Hepatol 2019;17:429–39.
5. Chen Z, Ren Y, Du XL, et al. Incidence and survival differences in esophageal cancer among ethnic groups in the United States. Oncotarget 2017;8:47037–51.
6. Then EO, Lopez M, Saleem S, et al. Esophageal cancer: An updated Surveillance Epidemiology and End Results database analysis. World J Oncol 2020;11:55–64.
7. Bus P, Lemmens VE, van Oijen MG, et al. Prognostic factors for medium- and long-term survival of esophageal cancer patient in The Netherlands. J Surg Oncol 2014;109:465–71.
8. Ellison LF. Differences in cancer survival in Canada by sex. Health Rep 2016;27:19–27.
9. Kauppila JH, Mattsson F, Brusselaers N, et al. Prognosis of oesophageal adenocarcinoma and squamous call carcinoma following surgery and no surgery in a nationwide Swedish cohort study. BMJ Open 2018;8:e021495.
10. Kauppila JH, Wahlin K, Lagergren P, et al. Sex differences in the prognosis after surgery for esophageal squamous cell carcinoma and adenocarcinoma. Int J Cancer 2019;144:1284–91.
11. Xie SH, Lagergren J. Risk factors for oesophageal cancer. Best Practice Res Clin Gastroenterol 2018;36–37:3–8.
12. Fabregat A, Sidiropoulos K, Garapati P, et al. The reactome pathway knowledgebase. Nucleic Acids Res 2016;44:D481–7.
13. Bohanes P, Yang D, Chhibar RS, et al. Influence of sex on the survival of patients with esophageal cancer. J Clin Oncol 2012;30:2265–72.
14. Huang T, Zhou Y, Cheng AS, et al. NOTCH receptors in gastric and other gastrointestinal cancers: Oncogenes or tumor suppressors? Mol Cancer 2016;15:80.
15. Wang LD, Zhou FY, Li XM, et al. Genome-wide association study of esophageal squamous cell carcinoma in Chinese subjects identifies a susceptibility locus at PLCE1. Nat Genet 2010;42:759–65.
16. Qiu M, Yang S, Wang M, et al. Prognostic evaluation of esophageal cancer patients with stages I-III. Aging (Albany NY) 2020;12:14736–53.
17. Minks J, Robinson WP, Brown CJ. A skewed view of X chromosome inactivation. J Clin Invest 2008;118:20–3.
18. Chaligné R, Heard E. X-chromosome inactivation in development and cancer. FEBS Lett 2014;588:2514–22.
19. Dunford A, Weinstock DM, Savova V, et al. Tumor-suppressor genes that escape from X- inactivation contribute to cancer sex bias. Nat Genet 2017; 49:10–6.
20. Migeon BR. The role of X inactivation and cellular mosaicism in women's health and sex- specific diseases. J Am Med Assoc 2006;295:1428–33.
21. Johnston CM, Lovell FL, Leongamornlert DA, et al. Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. PLoS Genet 2008;4:0088–98.
22. Libert C, Dejager L, Pinheiro I. The X chromosome in immune functions: When a chromosome makes the difference. Nat Rev Immunol 2010;10: 594–604.
23. Berletch JB, Disteche CM. Genes the escape X inactivation. Hum Genet 2012;130:237–45.
24. Botlagunta M, Vesuna F, Mironchik Y, et al. Oncogenic role of DDX3 in breast cancer biogenesis. Oncogene 2008;27:3912–22.
25. Yan X, Yan L, Su Z, et al. Zinc-finger protein X-linked is a novel predictor of prognosis in patients with colorectal cancer. Int J Clin Exp Pathol 2014; 7:3150–7.
26. Johansson C, Velupillai S, Tumber A, et al. Structural analysis of human KDM5B guides histone demethylase inhibitor development. Nat Chem Biol 2016;12:539–45.
27. Zhang Y, Duan C, Yang J, et al. Deubiquitinating enzyme USP9X regulates cellular clock function by modulating the ubiquitination and degradation of a core circadian protein BMAL1. Biochem J 2018;475:1507–22.
28. Ma H, Yang F, Lian M, et al. Dysregulation of zinc finger protein, X-linked (ZFX) impairs cell proliferation and induces apoptosis in human oral squamous cell carcinorma. Tumor Biol 2015;36:6103–12.
29. Wang C, Fu SY, da Wang M, et al. Zinc finger protein X-linked promotes expansion of EpCAM+ cancer stem-like cells in hepatocellular carcinoma. Mol Oncol 2017;11:455–69.
30. Kunze B, Wein F, Fang HY, et al. Notch signaling mediates differentiation in Barrett's esophagus and promotes progression to adenocarcinoma. Gastroenterology 2020;159:575–90.