

# Gene Regulatory Network Analysis Identifies Sex-Linked Differences in Colon Cancer Drug Metabolism



Camila M. Lopes-Ramos<sup>1,2</sup>, Marieke L. Kuijjer<sup>1,2</sup>, Shuji Ogino<sup>3,4,5,6</sup>, Charles S. Fuchs<sup>7,8,9</sup>, Dawn L. DeMeo<sup>10,11</sup>, Kimberly Glass<sup>10</sup>, and John Quackenbush<sup>1,2,12</sup>

## Abstract

Understanding sex differences in colon cancer is essential to advance disease prevention, diagnosis, and treatment. **Males have a higher risk of developing colon cancer and a lower survival rate than women. However, the molecular features that drive these sex differences are poorly understood.** In this study, we use both transcript-based and gene regulatory network methods to analyze RNA-seq data from The Cancer Genome Atlas for 445 patients with colon cancer. We compared gene expression between tumors in men and women and observed significant sex differences in sex chromosome genes only. We then inferred patient-specific gene regulatory networks and found significant regulatory differences between males and females, with drug and xenobiotics metabolism via cytochrome P450 pathways more strongly targeted in females. This finding was validated in a dataset of 1,193 patients from five independent studies. While targeting, the drug metabolism pathway did not change overall survival for males treated with adjuvant chemotherapy,

females with greater targeting showed an increase in 10-year overall survival probability, 89% [95% confidence interval (CI), 78–100] survival compared with 61% (95% CI, 45–82) for women with lower targeting, respectively ( $P = 0.034$ ). Our network analysis uncovers patterns of transcriptional regulation that differentiate male and female colon cancer and identifies differences in regulatory processes involving the drug metabolism pathway associated with survival in women who receive adjuvant chemotherapy. This approach can be used to investigate the molecular features that drive sex differences in other cancers and complex diseases.

**Significance:** A network-based approach reveals that sex-specific patterns of gene targeting by transcriptional regulators are associated with survival outcome in colon cancer. This approach can be used to understand how sex influences progression and response to therapies in other cancers. *Cancer Res*; 78(19): 5538–47. ©2018 AACR.

## Introduction

Significant differences between the sexes are observed during the development and progression of diseases, influencing disease incidence and survival. For many cancer types, such as colon, skin,

head and neck, esophagus, lung, and liver, males have a higher risk and higher mortality rates than females (1). Even though the higher risk in males might be attributed partially to occupational exposures and/or behavioral factors, such as diet, smoking, and alcohol consumption, after adjusting for these risk factors males still have a higher cancer risk, although residual confounding cannot be excluded (1–3). In colon cancer, females not only have reduced risk relative to males, but also have a better prognosis (4–6). Furthermore, females have a higher survival benefit from 5-fluorouracil (5-FU)-based adjuvant chemotherapy as compared with males (7). Pharmacokinetics also vary between the sexes; females experience greater toxicity from certain chemotherapies, including 5-FU, consistent with the lower 5-FU clearance observed in females (8, 9).

Sex differences in colon cancer have been largely attributed to sex hormones, yet the molecular mechanisms have not been established and clinical studies are contradictory (9). In general, studies point to the protective role of female hormones (estrogen) during colon cancer development and to the increased risk associated with male hormones (testosterone; refs. 10–12). The circadian system might also explain the better prognosis in females, polymorphisms in the *CLOCK* sequence, and the expression levels of miRNAs regulating the clock-genes were associated with longer overall survival of females with metastatic colorectal cancer compared with males (13). Although previous studies focused on a few targeted genes, a systems-based analysis that integrates multi-omics data can

<sup>1</sup>Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, Massachusetts. <sup>2</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. <sup>3</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. <sup>4</sup>Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts. <sup>5</sup>Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, Massachusetts. <sup>6</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts. <sup>7</sup>Yale Cancer Center, New Haven, Connecticut. <sup>8</sup>Department of Medicine, Yale School of Medicine, New Haven, Connecticut. <sup>9</sup>Smilow Cancer Hospital, New Haven, Connecticut. <sup>10</sup>Channing Division of Network Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts. <sup>11</sup>Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts. <sup>12</sup>Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts.

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

**Corresponding Author:** John Quackenbush, Harvard T.H. Chan School of Public Health and Dana-Farber Cancer Institute, 655 Huntington Ave, Boston, MA 02115. Phone: 617-432-9028; Fax: 617-432-5619; E-mail: [johnq@hsph.harvard.edu](mailto:johnq@hsph.harvard.edu)

**doi:** 10.1158/0008-5472.CAN-18-0454

©2018 American Association for Cancer Research.

provide insights into sex-specific regulatory processes associated with clinical outcome.

Regulatory networks characterize the complex cellular processes defined by a combination of signaling pathways and cell-type specific regulators. Each phenotype is defined by a characteristic network, while differences in network structures can shed light upon the biological processes that distinguish phenotypes. For example, gene regulatory network analysis has uncovered regulatory differences between cell lines and their tissues of origin, and between cancer subtypes (14, 15). Network-modeling approaches have also been valuable in determining sex-specific regulatory features in healthy tissues and in disease (16, 17).

Although both the risk for and outcome of colon cancer are different between men and women, clinical management is sex independent. This may be because the molecular features that drive these sex differences are poorly understood. We used network-modeling approaches, PANDA (18) and LIONESS (18, 19), to infer colon cancer patient-specific gene regulatory networks. We compared the male and female networks to identify genes that were targeted by transcription factors (TF) in a sex-specific manner (Fig. 1). We found that genes involved in drug and xenobiotics metabolism via cytochrome P450 were more strongly targeted by regulatory TFs in females; these results were validated in an independent dataset. Moreover, greater regulatory targeting of the drug metabolism pathway was predictive of longer survival in women who received adjuvant chemotherapy, but not in men.

## Materials and Methods

### Discovery dataset

We downloaded level 3 RNASeqV2 and clinical data for colon cancer from The Cancer Genome Atlas (TCGA) on June 16, 2016 (<https://tcga-data.nci.nih.gov>). We kept only primary tumor sam-

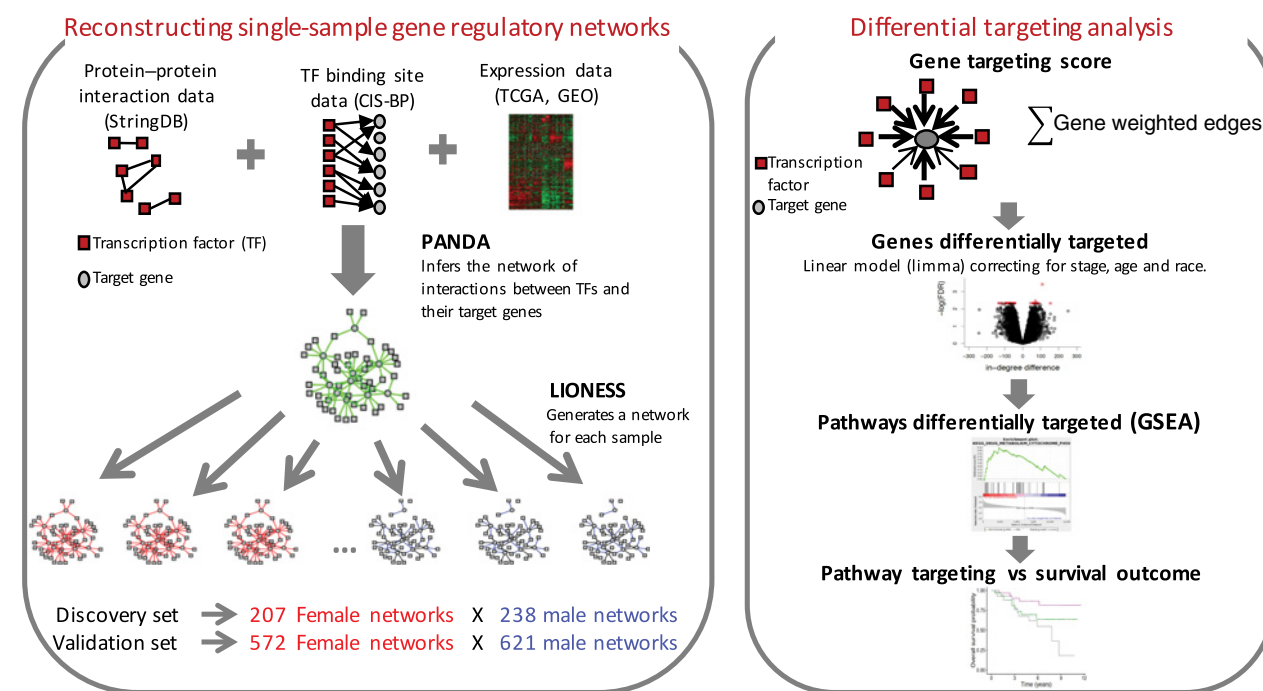
ples, removed samples that were not annotated for sex, and used principal component analysis (using the plotOrd function in metagenomeSeq 1.12.1; ref. 20) on genes located on the Y chromosome to identify and remove 9 potential sex-misannotated samples. After performing these quality control steps, the discovery dataset included 445 primary colon tumor samples before treatment, from 238 males and 207 females.

We filtered lowly expressed genes by removing genes with less than one count per million (CPM) in at least 104 samples, using the cpm function from R package edge R 3.18.1 (21), which corresponded to 5,571 of 20,249 genes. We chose 104 samples because that represents half of the samples in the smaller subgroup. To retain the same set of genes in the discovery and validation datasets, and the same set of genes for the differential expression, and differential targeting analysis, we kept only the genes overlapping the filtered genes in the discovery dataset, the validation dataset, and the genes in the TF/target gene regulatory prior used for creating the gene regulatory networks (see sections: "Validation dataset" and "Single-sample gene regulatory networks and differential targeting analysis"). This corresponded to 12,817 genes, which included genes on the sex chromosomes.

The expression data generated by TCGA were normalized using smooth quantile normalization (22), and batch was corrected for sequencing platforms (IlluminaGA and IlluminaHiSeq) and shipment date, as implemented in the R package qsmooth available on Github (<https://github.com/stephaniehicks/qsmooth>).

### Validation dataset

We searched the Gene Expression Omnibus (GEO) repository for colon cancer studies that included patient survival data and were obtained from the same microarray platform (Affymetrix Human Genome U133 Plus 2.0 Array). The validation dataset contained five independent studies: GSE14333, GSE17538,



**Figure 1.**

Study workflow. Overview of the approach used to reconstruct single-sample gene regulatory networks and to perform differential targeting analysis.

GSE33113, GSE37892, and GSE39582. Raw expression data and clinical data were downloaded from GEO using the R package GEOquery 2.36.0 on July 10, 2015. Raw expression data were preprocessed by frozen robust multiarray analysis (fRMA) using the R package fRMA 1.22.0 (23). Genes with multiple probe sets were represented by the probe set with the highest mean across all datasets. Then, we only kept the genes that overlapped with the discovery dataset and the genes in the TF/target gene regulatory prior. We removed samples obtained from rectal tumor, samples that were not annotated for sex, and potential sex-misannotated samples ( $n = 84$ ), as described previously. We performed batch correction by dataset series using the ComBat function implemented in the R package sva 3.18.0 (24). The final validation dataset was comprised of 1,193 primary colon tumor samples before treatment, from 621 males and 572 females, and 12,817 genes.

### Differential expression analysis

We used voom available in the R package limma 3.26.9 (25) to compare gene expression between colon tumor samples from males and females after adjusting for the covariates age, race, and disease stage by the Union for International Cancer Control (UICC). Voom models the mean–variance relationship of sequencing counts and incorporates this information into the limma empirical Bayes differential expression analysis. We performed multiple testing corrections using the method of Benjamini–Hochberg (26).

### Single-sample gene regulatory networks and differential targeting analysis

**Network reconstruction.** We used the PANDA (18) and LIONESS (19) algorithms to reconstruct gene regulatory networks for each sample in both the discovery and the validation datasets. The networks were inferred from three types of data: TF/target gene regulatory prior (obtained by mapping TF motifs from the Catalog of Inferred Sequence Binding Preferences (CIS-BP; ref. 27) to the promoter of their putative target genes), protein–protein interaction data (using the interaction scores from StringDb v1.0 (28) between all TFs in the regulatory prior), and gene expression (obtained from the discovery or validation datasets). The TF/target gene regulatory prior and the protein–protein interaction data were generated as described by Sonawane and colleagues (29), and then we kept only the genes that matched the genes expressed in the discovery and validation datasets. Our TF/target gene regulatory prior consisted of 661 TFs targeting 12,817 genes, and the protein–protein interaction data consisted of interactions between the 661 TFs.

PANDA begins with a prior network of TFs and their potential targets based on mapping TF motifs to the genome, and then combines with protein–protein interaction data and gene expression data. PANDA uses message passing to find agreement between these three data types and to optimize the structure of the network. LIONESS can be applied to PANDA-constructed networks to estimate sample-specific gene regulatory networks. LIONESS uses an iterative process that leaves out each individual in a population, estimates the network with and without that individual, and then interpolates edge weights to derive an estimate for the network active in that single individual.

**Network interpretation.** An edge (TF/target gene relationship) in a network modeled using PANDA and LIONESS represents the

likelihood that a gene is regulated by a particular TF. PANDA first transforms the TF/target gene regulatory prior network into Z-score space and then optimizes the initial edges using a message-passing approach. This approach aims to find agreement between the different data types, that is, how similar the targeting profile of a TF is to the coexpression of its target genes and whether TFs can form complexes to regulate a specific gene. Lower-valued edge weights indicate a small likelihood of regulation, whereas edge weights with higher values indicate an increased likelihood of regulation.

**Single-sample gene targeting score.** For each sample's gene regulatory network, we calculated each gene's targeting score, equivalent to the gene's in-degree (defined as the sum of the gene's incoming edge weights from all TFs in the network). Next, we compared the gene targeting score between males and females using a linear regression model and correcting for the covariates age, race, and disease stage, as available in the R package limma 3.26.9 (30).

**Network visualization.** The subnetwork was illustrated using Cytoscape default yFiles Organic layout (version 3.4.0; ref. 31), where each edge connects a TF to a target gene, and the color represents the average edge weight difference between the male and female networks.

### Pathway enrichment analysis

To perform pathway enrichment analysis, we used pre-ranked gene set enrichment analysis (GSEA; Java command line version 2-2.0.13; ref. 32), and the gene sets obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database (33) that were downloaded from the Molecular Signatures Database (MSigDB; <http://www.broadinstitute.org/gsea/msigdb/collections.jsp>; "c2.cp.kegg.v5.0.symbols.gmt"). We only considered gene sets of size greater than 15 and less than 500 after filtering out those genes not in the expression dataset, which restricted our analysis to 176 gene sets. To perform the analysis, all genes were ranked by the  $t$ -statistic produced by the voom differential expression analysis or by the limma differential targeting analysis after adjusting for covariates.

### Survival analysis

We performed Kaplan–Meier survival analysis as implemented in the R package survival 2.41–3, and the  $P$  values were computed using the log-rank test. The survival curves were plotted using the function ggsurv on the GGally package 1.3.2.

## Results

### Data features

We investigated the regulatory processes that drive sex differences in colon cancer by performing network analysis in a discovery dataset followed by a validation analysis in an independent dataset. For the discovery dataset, we analyzed RNA sequencing (RNA-seq) data of primary colon tumor samples from TCGA. We removed potential sex-misannotated samples after assessing the expression of Y chromosome genes and retained 445 samples, obtained before treatment, for the primary analysis.

For the validation dataset, we downloaded raw microarray data from five independent studies in GEO (34), which were profiled on the same microarray platform (Affymetrix Human Genome U133 Plus 2.0 Array). We corrected the data for study batch and removed potential sex-misannotated samples. The final

**Table 1.** Clinical features for the discovery (TCGA) and validation (GEO) datasets included in the study

	Discovery dataset			Validation dataset		
	Female	Male	P	Female	Male	P
Number of patients	207	238	—	572	621	—
Age	66.18 ± 14.22	67.53 ± 12.02	0.285	68.49 ± 13.69	65.10 ± 12.70	1.06E–05
Race						
Black	30	28	0.315	5	8	0.640
Caucasian	102	109		91	100	
Other	3	9		11	7	
Not specified	72	92		465	506	
Anatomic region						
Cecum	48	56	0.992	0	0	0.089
Ascending colon	39	47		192	186	
Hepatic flexure	13	13		0	0	
Transverse colon	17	18		0	0	
Splenic flexure	3	3		0	0	
Descending colon	8	12		221	279	
Sigmoid colon	70	77		0	0	
Rectosigmoid junction	1	0		0	0	
Not specified	8	12		159	156	
UICC stage						
1	34	41	0.864	26	31	0.872
2	81	92		225	249	
3	62	63		156	153	
4	26	35		54	59	
Not specified	4	7		111	129	

NOTE: *t* test was performed for age, and Fisher exact test for all other variables.

validation dataset included 1,193 primary colon tumor samples obtained before treatment. Patient clinical features are summarized in Table 1 and are extensively presented in Supplementary Table S1. The number of males and females were well matched, and for all downstream analysis we controlled for potential differences between the sexes for age, race, and colon cancer stage.

#### Autosomal genes are not differentially expressed between males and females in colon cancer

As a baseline for our network analysis, we first performed a differential expression comparison between males and females using voom (25). We found 12 genes significantly different between males and females with an absolute fold change greater than two and an FDR less than 0.1 (Fig. 2A). All the differentially expressed genes were located on the X and Y chromosomes, and no genes were found to be differentially expressed when we excluded sex chromosomes from the analysis.

We also tested whether there were changes in gene expression levels across the genome that were below our detection threshold but that might be relevant for sex differences in colon cancer. Therefore, to understand the biological functions associated with changes in gene expression, we ranked all genes based on their statistical significance (*t*-statistic) and performed preranked GSEA using KEGG pathways (32, 33). We did not find any significant differential enrichment of KEGG pathways between males and females. Thus, based only on gene expression differences between the sexes, it was not possible to identify biological pathways that distinguish males and females with colon cancer.

#### Differential targeting of biological pathways in males and females

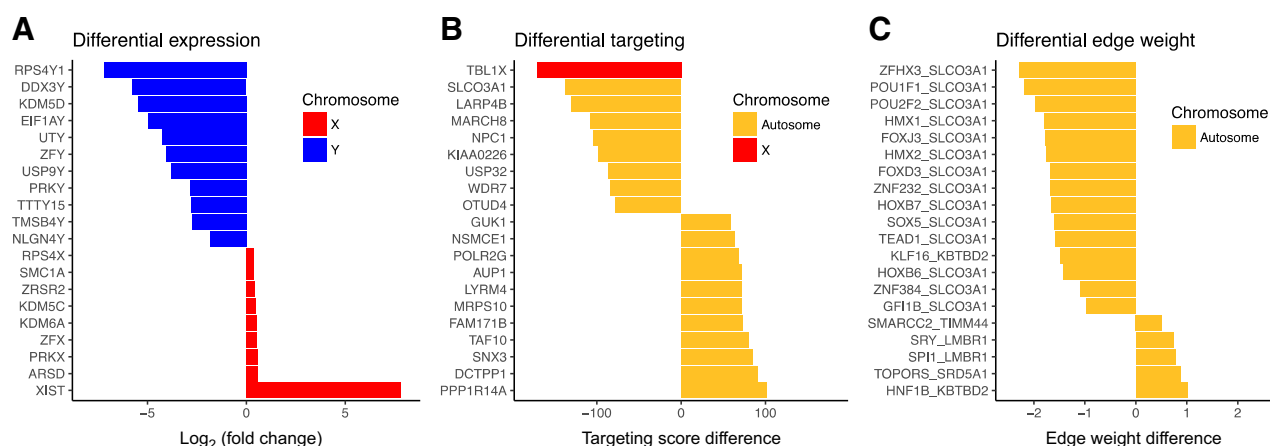
In previous studies, we have found gene regulatory network analysis provides greater insight into altered biological processes than do simple tests of differential gene expression. In particular, our work has shown that TFs often play specific regulatory roles through changes in their targeting patterns that are independent

of their mRNA expression level (14, 17, 29). Thus, we performed a gene regulatory network analysis using PANDA (18) and LIONESS (Fig. 1; ref. 19), inferring gene regulatory networks for each individual in our discovery population and independently for the validation dataset.

PANDA is a message-passing algorithm that starts with a TF/target gene prior regulatory network based on a motif scan that maps TF-binding sites to the promoter of their putative target genes, then integrates the prior regulatory network with protein–protein interaction data and gene expression data. In the gene regulatory network modeled with PANDA, each edge connects a TF to a target gene, and the associated edge weight reflects the strength of the inferred regulatory relationship. We initiated PANDA with TF-binding sites from CISBP (27), protein–protein interaction data from StringDb (28), and expression data from either the discovery or the validation dataset to estimate aggregate gene regulatory networks based on all samples. Then, we used LIONESS to estimate each sample-specific gene regulatory network. This method enables the identification of individual-specific regulatory networks, allowing us to associate network properties with clinical information.

As a control for the differential targeting analysis, we started by comparing colon cancer with healthy colon tissue. For the healthy colon networks, we used the single-sample gene regulatory networks reconstructed by Chen and colleagues (16), which were modeled on data from the Genotype-Tissue Expression (GTEx) project (35). For the colon cancer networks, we reconstructed gene regulatory networks for each sample, as shown in Fig. 1. Next, we performed differential targeting analysis. For each gene, we calculated a gene targeting score equal to the sum of all incoming edge weights a gene receives from all TFs in the network (its "in-degree"). We then used a linear regression model to test whether there is a significant difference in the gene targeting score between healthy and cancer tissues using limma (24), and adjusting for age and race. All genes were ranked by their statistical significance, and we used preranked GSEA to identify the pathways enriched for the



**Figure 2.**

Differential expression and differential targeting between male and female colon cancer. **A**, Expression log<sub>2</sub> (fold change) of the top 20 differentially expressed genes between males and females in the discovery dataset. **B**, Gene targeting score difference of the top 20 genes differentially targeted between males and females in the discovery dataset. **C**, Edge weight difference of the top 20 edges most significantly different between males and females in the discovery dataset. Positive values indicate higher levels in females, and negative values indicate higher levels in males.

differentially targeted genes. As expected, when comparing healthy and cancer tissues, we found that pathways enriched for genes more strongly targeted in cancer were associated with apoptosis and immune signaling, whereas pathways enriched for genes more strongly targeted in healthy tissues were associated with focal adhesion and extracellular matrix interaction (Supplementary Fig. S1). Similar pathways were found to be differentially targeted between healthy and cancer tissues independent of the sex analyzed.

To explore sex differences in colon cancer, we performed a similar differential targeting analysis to compare the male colon cancer networks with the female colon cancer networks and identified the genes differentially targeted between the sexes. We found 21 genes differentially targeted between males and females (FDR < 0.1); only one of the genes (*TBL1X*) is located on sex chromosomes (Fig. 2B). This stands in contrast to the differential expression analysis, which found genes exclusively on the sex chromosomes. Although the differential expression analysis gives the expression difference for each individual gene, the differential targeting analysis allows us to infer how much a gene is targeted by a set of TFs and may better elucidate the biological differences in colon cancer between males and females.

We also investigated the most significantly different edges between males and females (Fig. 2C). For this, we used limma to compare the edge weights connecting TFs with genes between males and females (correcting for age, race, and stage). Consistent with the higher targeting score of *SLCO3A1* in male networks (Fig. 2B), we found the most significantly different edges include many TFs targeting *SLCO3A1* (Fig. 2C).

We ranked all genes based on their differential targeting statistical significance (*t*-statistic), and performed preranked GSEA. We found that genes more strongly targeted in males were enriched for NOTCH signaling pathway (FDR = 0.002), mTOR signaling pathway (FDR = 0.01), and WNT signaling pathway (FDR = 0.02). These pathways have key roles in colon cancer biology (36, 37) and their higher targeting in males may be related with the higher risk and lower survival rates observed in colon cancer in males. In females, we found significantly higher target-

ing of pathways associated with metabolism, including steroid hormone biosynthesis (FDR = 0.02), metabolism of xenobiotics by cytochrome P450 (FDR = 0.02), and drug metabolism-cytochrome P450 (FDR = 0.04; Table 2). This indicates that regulatory differences between the sexes in the drug metabolism pathway may impact the response to chemotherapy treatment and survival outcome in a sex-specific manner. We therefore further investigated this pathway.

#### Genes involved in drug metabolism are more strongly targeted by TFs in females

The network analysis identified significant sex differences in the regulatory pattern of the drug metabolism pathway. Because the regulatory differences are present in nontreated tumor tissues, these differences can set the basis on how each sex will respond to chemotherapy treatment and impact sex-specific management and outcome.

We compared the edge weights connecting TFs with genes using limma (correcting for age, race, and stage). Figure 3A illustrates the edges around genes in the drug metabolism pathway that are most significantly different between males and females. Of the 2,830 edges around genes in the drug metabolism pathway, 220 edges have an FDR < 0.05 and are represented in Fig. 3A. Overall, we observe most significant edges are those with strong regulatory targeting of genes in the female networks, contributing to the higher gene targeting score we had found. Many of the genes more strongly targeted in females, such as *GSTO1*, *GSTA4*, *GSTT2*, *MGST2*, *MGST3*, belong to the family of glutathione S-transferases, which are important in the detoxification leading to the elimination of toxic compounds (38), whereas in the male networks only *GSTM2* is more strongly targeted. We also find that some genes, such as *CYP2D6* and *GSTM4*, have similar overall gene targeting score, while being targeted by different sets of TFs in male and female networks. For example, in the female networks, *CYP2D6* is more strongly targeted by TFs responsive to estrogens, such as estrogen-related receptor  $\alpha$  (*ESRRA*), estrogen-related receptor  $\beta$  (*ESRRB*), and estrogen-related receptor  $\gamma$  (*ESRRG*). In the male networks, *CYP2D6* is more strongly targeted by SRY-box

**Table 2.** Pathways differentially targeted between male and female colon cancer regulatory networks

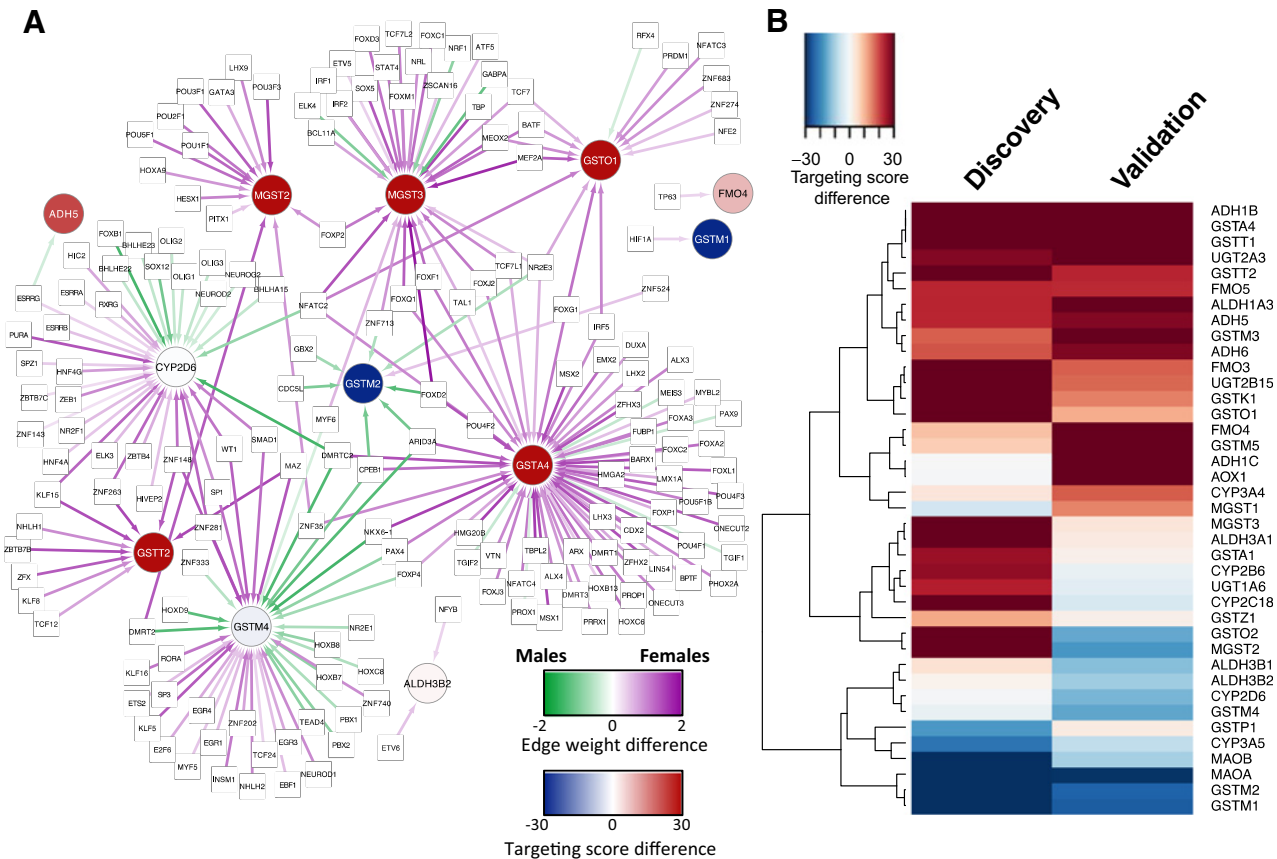
Enrichment in males			Enrichment in females		
Pathway	NES	FDR	Pathway	NES	FDR
Acute myeloid leukemia	-2.207	0.000	Oxidative phosphorylation	2.637	0.000
Endometrial cancer	-2.230	0.000	Parkinson disease	2.485	0.000
Chronic myeloid leukemia	-2.157	0.001	Ribosome	2.287	0.000
Notch signaling pathway	-2.114	0.002	Alzheimer disease	1.957	0.003
Phosphatidylinositol signaling system	-2.062	0.002	Proteasome	1.968	0.004
ErbB signaling pathway	-2.006	0.002	Peroxisome	1.914	0.006
Pancreatic cancer	-2.012	0.003	Huntington disease	1.872	0.008
Focal adhesion	-2.018	0.003	Terpenoid backbone biosynthesis	1.807	0.015
Colorectal cancer	-2.026	0.003	Metabolism of xenobiotics by cytochrome p450	1.791	0.015
Prostate cancer	-1.962	0.003	Steroid hormone biosynthesis	1.761	0.018
Jak stat signaling pathway	-1.923	0.004	Protein export	1.734	0.021
Adherens junction	-1.914	0.004	Histidine metabolism	1.736	0.022
Arrhythmogenic right ventricular cardiomyopathy arvc	-1.917	0.004	Amyotrophic lateral sclerosis als	1.677	0.034
Fc gamma r mediated phagocytosis	-1.886	0.005	Drug metabolism cytochrome p450	1.648	0.041
Chemokine signaling pathway	-1.849	0.007	Tyrosine metabolism	1.590	0.064

Abbreviation: NES, normalized enrichment score.

12 (SOX12), which belongs to a family of TFs characterized by the presence of a DNA-binding high-mobility group domain, homologous to that of sex-determining region Y (SRY; ref. 39).

Next, we confirmed the drug metabolism pathway differential targeting in an independent dataset. As described previously, our validation dataset was obtained from GEO and included 1,193

samples from five independent studies. We repeated the pipeline described above to reconstruct single-sample gene regulatory networks and to perform differential targeting analysis in the validation dataset. The pathways more strongly targeted in males did not reach statistical significance in this validation dataset (Supplementary Table S2). However, we again identified the



**Figure 3.** Differential targeting of genes in the drug metabolism pathway. **A**, Subnetwork representation of the edges in the drug metabolism pathway that are most significantly different between male and female regulatory networks (FDR < 0.05). **B**, Heatmap of gene targeting score difference for all the analyzed genes in the drug metabolism-cytochrome P450 pathway for the discovery and validation dataset analyses. For visualization purposes, the color scale was saturated at 30. Positive values indicate higher levels in females, and negative values indicate higher levels in males.

drug metabolism-cytochrome P450 (FDR = 0.009) and the metabolism of xenobiotics by cytochrome P450 (FDR = 0.012) pathways as enriched for genes that were more strongly targeted in females.

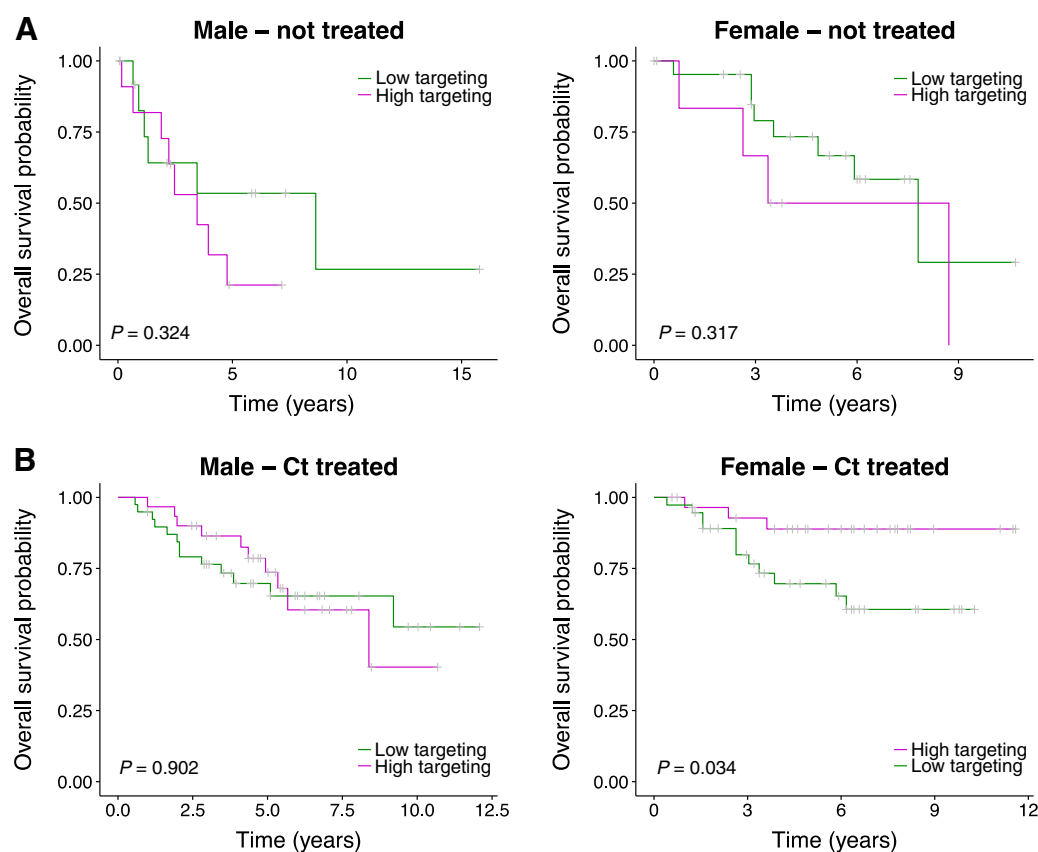
Some known tumor molecular features, such as CpG island methylator phenotype (CIMP), chromosomal instability (CIN) phenotype, and mismatch repair (MMR) deficiency, may be sex biased and/or associated with disease prognosis and treatment response. Because part of the validation group was profiled for these features, we repeated the differential targeting analysis for the validation group adjusting for CIMP, CIN, and MMR. Again, we confirmed females have higher targeting of the drug metabolism-cytochrome P450 (FDR = 0.032) and the metabolism of xenobiotics by cytochrome P450 (FDR = 0.038).

This validation in an independent dataset, and one derived using a different technology (microarrays rather than RNA-seq), gives us a high degree of confidence in the observed sex-specific regulation of drug metabolism. Indeed, the genes in the drug metabolism pathway with higher gene targeting score difference are highly consistent between the discovery and validation datasets (Fig. 3B). The main regulatory sex differences are associated with genes important during catabolism and detoxification, for example, *ADH1B*, *GSTA4*, and *FMO5*.

We also analyzed differential targeting of the drug metabolism pathway between males and females in healthy colon tissue. We compared healthy colon tissue samples from 223 males and 153 females (obtained from Chen and colleagues; ref. 22) by performing differential targeting analysis after adjusting for age and race. We found no significant difference for the drug metabolism-cytochrome P450 pathway (FDR = 0.31) nor in the metabolism of xenobiotics by cytochrome P450 (FDR = 0.44) in healthy tissues.

#### Higher targeting of the drug metabolism pathway is associated with better overall survival in females treated with adjuvant chemotherapy

Considering that treatment benefit and survival outcome are largely dependent on tumor stage, we performed a survival analysis to evaluate chemotherapy treatment benefit in each stage. Only patients with stage and treatment information were selected for these survival analyses ( $n = 514$  from GSE39582). Stage I patients were excluded because none of these patients were treated with chemotherapy. For stage II, there was no overall survival difference between patients with and without chemotherapy treatment (Supplementary Fig. S2). This was expected, as many studies show only a small or no benefit



**Figure 4.**

Overall survival analysis based on targeting of the drug metabolism pathway. Kaplan-Meier curve of patients with stage III colon cancer not treated (A) or treated (B) with adjuvant chemotherapy. Patients were stratified by low- and high-targeting groups based on the median gene targeting score across all genes in the drug metabolism pathway.  $P$  values were computed using the log-rank test to evaluate the overall survival risk between the low targeting (green) and high targeting (purple) subgroups of patients.

from chemotherapy for stage II patients, and only a small subgroup of patients with more aggressive tumors benefits from treatment (40). Chemotherapy treatment benefit was clear for stage III patients: Treated patients had a significantly higher overall survival probability than nontreated patients. For stage IV, we did not find a survival difference after chemotherapy treatment, possibly due to the small number of samples included in this analysis ( $n = 45$ ). Therefore, we limited subsequent survival analysis to stage III patients ( $n = 190$ ).

To evaluate whether the strength of targeting for drug metabolism pathway genes is associated with outcome, we performed overall survival analysis after stratifying patients by how strongly the pathway was targeted. For this, patients in the validation dataset were divided in two groups of equal size (named low- and high targeting groups) based on the median targeting score of genes in the drug metabolism pathway. For patients not treated with adjuvant chemotherapy, targeting of the drug metabolism pathway did not change the overall survival probability, showing that high targeting is not a prognostic factor (Fig. 4A).

Next, we assessed the clinical impact of the drug metabolism pathway targeting for patients treated with adjuvant chemotherapy. Consistent with the weaker targeting of the drug metabolism pathway observed in male networks, we found that gene targeting strength of the drug metabolism pathway did not change the overall survival probability for males receiving adjuvant chemotherapy (Fig. 4B). In contrast, in females who received treatment, higher targeting of the drug metabolism pathway was associated with a better overall survival. The 10-year overall survival probability was 61% [95% confidence interval (CI), 45–82] for females with low targeting and 89% (95% CI, 78–100) for females with high targeting (log-rank test  $P = 0.034$ ). There was no statistically significant difference between the females with high- and low targeting groups for the available covariates (age, MMR, CIMP, and CIN).

## Discussion

The use of standard differential expression analysis between males and females has found the expected differential regulation of sex chromosome genes (41–43). However, differential targeting analysis identified sex-specific differences are also driven by global transcriptional regulatory processes, a phenomenon we had seen previously in chronic obstructive pulmonary disease (17). Using a network-based approach, we found that not only are genes associated with drug metabolism differentially regulated in males and females, but also that patterns of gene targeting are associated with clinical outcome, particularly in women. Specifically, we find that greater targeting of the drug metabolism pathway is associated with a better overall survival in females treated with adjuvant chemotherapy.

Many of the genes associated with the drug metabolism pathway are highly polymorphic, and they have been associated with colon cancer risk (44–46). For example, a meta-analysis showed that *GSTM1* and *GSTT1*-null genotypes increase the risk of colorectal cancer in Caucasian populations (47). For lymphoblastoid cell lines (LCL), it has been shown that the KEGG pathway "metabolism of xenobiotics by cytochrome P450" is enriched for genes more expressed in females compared with males (48). In healthy colon tissues, we found no significant sex differences in the regulatory patterns of the drug metabolism pathway. The

differences were evident in colon cancer tissues, and these differences were independent of the disease stage.

The genes with the largest sex differences belong to the glutathione S-transferase (GST) family and they are highly targeted in females. GSTs are metabolizing enzymes that play a key role during neutralization and elimination of toxic compounds and xenobiotics (38). Thus, considering the role of the differentially targeted genes found on samples before treatment, one can expect that each sex might respond differently to chemotherapeutic treatment and ultimately have a different survival outcome. We found that stronger regulation of genes in the drug metabolism pathway in pretreatment samples is associated with better outcome in women treated with chemotherapy. This stronger regulation of the pathway in pretreated tumors might set the baseline for how the regulators and target genes will respond to therapy. Tumors with a stronger control of the pathway may respond better to treatment, whereas tumors without this strong regulatory control may exhibit a weak response to chemotherapy. Following treatment, we would expect to see differences in the expression of drug metabolism genes between males and females. This could be further investigated if we had available samples from posttreatment biopsies.

Changes in the regulatory network might come from subtle changes in gene expression including patient-specific changes in different genes affecting the same pathway. In addition, changes in the regulatory network need not only come from differences in TF expression levels. For example, in modeling gene regulatory networks in 38 tissues, we found that TF expression levels are often independent of their regulatory potential across many human tissues (29). Instead, tissue-specific expression is associated with TFs that change their targeting patterns to activate tissue-specific regulatory roles. Differences in TF regulation can be caused by the protein abundance of a TF, the residency time of a TF on the DNA, the abundance levels of other TFs competing for binding to a particular motif, epigenetic modifications that interfere with transcription, posttranscriptional regulation, as well as other mechanisms (49, 50).

We note the analyzed data may be influenced by heterogeneity (cellular or clinical) and risk factor profiles (such as dietary habits, and family history). To reduce confounding effects, the data were corrected for known covariates, such as age, race, and disease stage. Although our study has a small number of samples treated with adjuvant chemotherapy and with survival information, the identification of differential regulation of drug metabolism pathways in independent datasets provides support for our conclusions. This suggests that clinical trials and other experiments should carefully consider the manifestation of sex differences and be statistically powered to understand the impact of sex in outcomes.

Even though there are significant sex differences regarding risk, prognosis, treatment response, and chemotherapeutic toxicity related to colon cancer, management of colon cancer is not based on sex, and the molecular features that drive the sex differences are still poorly understood. We found that colon cancer has significant sex differences related to TF regulatory patterns. Most importantly, targeting of the drug metabolism pathway was predictive of higher overall survival in women who received adjuvant chemotherapy. The determination that sex-specific targeting could discriminate between long-term and short-term survivors raises the possibility of using gene regulatory network analyses in other diseases. Indeed, the regulatory network analysis



method described here can easily be used to understand how sex influences progression and response to therapies in other cancer types and complex diseases and may help motivate development of sex-specific approaches to disease treatment.

### Disclosure of Potential Conflicts of Interest

J. Quackenbush is a consultant/advisory board member for Caris Life Sciences and Merck KGaA. No potential conflicts of interest were disclosed by the other authors.

### Authors' Contributions

**Conception and design:** C.M. Lopes-Ramos, M.L. Kuijjer, K. Glass, J. Quackenbush

**Development of methodology:** C.M. Lopes-Ramos, M.L. Kuijjer, J. Quackenbush

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** M.L. Kuijjer, C.S. Fuchs

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** C.M. Lopes-Ramos, M.L. Kuijjer, S. Ogino, C.S. Fuchs, D.L. DeMeo, J. Quackenbush

**Writing, review, and/or revision of the manuscript:** C.M. Lopes-Ramos, M.L. Kuijjer, S. Ogino, D.L. DeMeo, K. Glass, J. Quackenbush

### References

- Clocchiatti A, Cora E, Zhang Y, Dotto GP. Sexual dimorphism in cancer. *Nat Rev Cancer* 2016;16:330–9.
- Wei EK, Giovannucci E, Wu K, Rosner B, Fuchs CS, Willett WC, et al. Comparison of risk factors for colon and rectal cancer. *Int J Cancer* 2004;108:433–42.
- Edgren G, Liang L, Adami H-O, Chang ET. Enigmatic sex disparities in cancer incidence. *Eur J Epidemiol* 2012;27:187–96.
- Koo JH, Jalaludin B, Wong SKC, Kneebone A, Connor SJ, Leong RWL. Improved survival in young women with colorectal cancer. *Am J Gastroenterol* 2008;103:1488–95.
- Majek O, Gondos A, Jansen L, Emrich K, Holleczer B, Katalinic A, et al. Sex differences in colorectal cancer survival: population-based analysis of 164,996 colorectal cancer patients in Germany. Suzuki H, editor. *PLoS One* 2013;8:e68077.
- Quirt JS, Nanji S, Wei X, Flemming JA, Booth CM. Is there a sex effect in colon cancer? Disease characteristics, management, and outcomes in routine clinical practice. *Curr Oncol* 2017;24:e15–23.
- Elsaleh H, Joseph D, Grief F, Zeps N, Spry N, Iacopetta B. Association of tumour site and sex with survival benefit from adjuvant chemotherapy in colorectal cancer. *Lancet* 2000;355:1745–50.
- Wang J, Huang Y. Pharmacogenomics of sex difference in chemotherapeutic toxicity. *Curr Drug Discov Technol* 2007;4:59–68.
- Barzi A, Lenz AM, Labonte MJ, Lenz H-J. Molecular pathways: estrogen pathway in colorectal cancer. *Clin Cancer Res* 2013;19:5842–8.
- Chlebowski RT, Wactawski-Wende J, Ritenbaugh C, Hubbell FA, Ascensao J, Rodabough RJ, et al. Estrogen plus progestin and colorectal cancer in postmenopausal women. *N Engl J Med* 2004;350:991–1004.
- Lin JH, Zhang SM, Rexrode KM, Manson JE, Chan AT, Wu K, et al. Association between sex hormones and colorectal cancer risk in men and women. *Clin Gastroenterol Hepatol* 2013;11:419–424.e1.
- Amos-Landgraf JM, Heijmans J, Wielenga MCB, Dunkin E, Krentz KJ, Clifton L, et al. Sex disparity in colonic adenomagenesis involves promotion by male hormones, not protection by female hormones. *Proc Natl Acad Sci U S A* 2014;111:16514–9.
- Garufi C, Giacomini E, Torsello A, Sperduti I, Melucci E, Mottolise M, et al. Gender effects of single nucleotide polymorphisms and miRNAs targeting clock-genes in metastatic colorectal cancer patients (mCRC). *Sci Rep* 2016;6:34006.
- Lopes-Ramos CM, Paulson JN, Chen C-Y, Kuijjer ML, Fagny M, Platig J, et al. Regulatory network changes between cell lines and their tissues of origin. *BMC Genomics* 2017;18:723.
- Glass K, Quackenbush J, Spentzos D, Haibe-Kains B, Yuan G-C. A network model for angiogenesis in ovarian cancer. *BMC Bioinformatics* 2015;16:1–17.
- Chen C-Y, Lopes-Ramos CM, Kuijjer ML, Paulson JN, Sonawane AR, Fagny M, et al. Sexual dimorphism in gene expression and regulatory networks across human tissues. *bioRxiv* 2016. doi: <https://doi.org/10.1101/082289>.
- Glass K, Quackenbush J, Silverman EK, Celli B, Rennard SI, Yuan G-C, et al. Sexually-dimorphic targeting of functionally-related genes in COPD. *BMC Syst Biol* 2014;8:1–17.
- Glass K, Huttenhower C, Quackenbush J, Yuan G-C. Passing messages between biological networks to refine predicted interactions. *PLoS One* 2013;8:e64832.
- Kuijjer ML, Tung M, Yuan G-C, Quackenbush J, Glass K. Estimating sample-specific regulatory networks. *arXiv* 2018. arXiv:1505.06440v2 [q-bio.MN].
- Paulson J, Talukder H, Pop M, Bravo H. metagenomeSeq: Statistical analysis for sparse high-throughput sequencing. Available from: <https://www.bioconductor.org/packages/devel/bioc/vignettes/metagenomeSeq/inst/doc/metagenomeSeq.pdf>.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139–40.
- Hicks SC, Okrah K, Paulson JN, Quackenbush J, Irizarry RA, Bravo HC. Smooth quantile normalization. *Biostatistics* 2017;56:222–9.
- McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). *Biostatistics* 2010;11:242–53.
- Leek JT, Storey JD, Kruglyak L, Weinblatt M, SN A. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 2007;3:e161.
- Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014;15:R29.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 1995;57:289–300.
- Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 2014;158:1431–43.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015;43:D447–52.
- Sonawane AR, Platig J, Fagny M, Chen C-Y, Paulson JN, Lopes-Ramos CM, et al. Understanding tissue-specific gene regulation. *Cell Rep* 2017;21:1077–88.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.

**Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases):** C.S. Fuchs, J. Quackenbush  
**Study supervision:** K. Glass, J. Quackenbush

### Acknowledgments

M.L. Kuijjer is supported by grants from the Charles A. King Trust Postdoctoral Research Fellowship Program, Sara Elizabeth O'Brien Trust, Bank of America, N.A., Co-Trustees and from the NCI at the NIH through P50CA165962. S. Ogino is supported by the NIH/NCI through R35CA197735. D.L. DeMeo is supported by grants from the National Heart, Lung, and Blood Institute (NHLBI) at the NIH, including P01 HL105339, P01 HL114501, P01 HL132825. K. Glass is supported by the NIH/NHLBI through K25HL133599. J. Quackenbush is supported by grants from the NIH/NCI, including 5P50CA127003, 1R35CA197449, and 1R01CA205406.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 22, 2018; revised June 4, 2018; accepted July 20, 2018; published first October 1, 2018.

31. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
32. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
33. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2015;44:D457–62.
34. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res* 2011;39:D1005–10.
35. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. *Nat Genet* 2013;45:580–5.
36. Capaccione KM, Pine SR. The Notch signaling pathway as a mediator of tumor survival. *Carcinogenesis* 2013;34:1420–30.
37. Zhan T, Rindtorff N, Boutros M. Wnt signaling in cancer. *Oncogene* 2017;36:1461–73.
38. Townsend DM, Tew KD. The role of glutathione-S-transferase in anti-cancer drug resistance. *Oncogene* 2003;22:7369–75.
39. Jiang T, Hou C-C, She Z-Y, Yang W-X. The SOX gene family: function and regulation in testis determination and male fertility maintenance. *Mol Biol Rep* 2013;40:2187–94.
40. Benson AB, Schrag D, Somerfield MR, Cohen AM, Figueredo AT, Flynn PJ, et al. American society of clinical oncology recommendations on adjuvant chemotherapy for stage II colon cancer. *J Clin Oncol* 2004;22:3408–19.
41. Ma J, Malladi S, Beck AH. Systematic analysis of sex-linked molecular alterations and therapies in cancer. *Sci Rep* 2016;6:19119.
42. Vawter MP, Evans S, Choudary P, Tomita H, Meador-Woodruff J, Molnar M, et al. Gender-specific gene expression in post-mortem human brain: localization to sex chromosomes. *Neuropsychopharmacology* 2004;29:373–84.
43. Werner RJ, Schultz BM, Huhn JM, Jelinek J, Madzo J, Engel N. Sex chromosomes drive gene expression and regulatory dimorphisms in mouse embryonic stem cells. *Biol Sex Differ* 2017;8:28.
44. van der Logt EMJ, Bergevoet SM, Roelofs HMJ, van Hooijdonk Z, te Morsche RHM, Wobbes T, et al. Genetic polymorphisms in UDP-glucuronosyltransferases and glutathione S-transferases and colorectal cancer risk. *Carcinogenesis* 2004;25:2407–15.
45. Hlavata I, Vrana D, Smerhovský Z, Pardini B, Naccarati A, Vodicka P, et al. Association between exposure-relevant polymorphisms in CYP1B1, EPHX1, NQO1, GSTM1, GSTP1 and GSTT1 and risk of colorectal cancer in a Czech population. *Oncol Rep* 2010;24:1347–53.
46. García-González MA, Quintero E, Bujanda L, Nicolás D, Benito R, Strunk M, et al. Relevance of GSTM1, GSTT1, and GSTP1 gene polymorphisms to gastric cancer susceptibility and phenotype. *Mutagenesis* 2012;27:771–7.
47. Economopoulos KP, Sergentanis TN. GSTM1, GSTT1, GSTP1, GSTA1 and colorectal cancer risk: a comprehensive meta-analysis. *Eur J Cancer* 2010;46:1617–31.
48. Shen JJ, Wang T-Y, Yang W. Regulatory and evolutionary signatures of sex-biased genes on both the X chromosome and the autosomes. *Biol Sex Differ* 2017;8:35.
49. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The human transcription factors. *Cell* 2018;172:650–65.
50. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 2017;356:eaaj2239.

# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## Gene Regulatory Network Analysis Identifies Sex-Linked Differences in Colon Cancer Drug Metabolism

Camila M. Lopes-Ramos, Marieke L. Kuijjer, Shuji Ogino, et al.

*Cancer Res* 2018;78:5538-5547.

**Updated version** Access the most recent version of this article at:  
<http://cancerres.aacrjournals.org/content/78/19/5538>

**Cited articles** This article cites 47 articles, 6 of which you can access for free at:  
<http://cancerres.aacrjournals.org/content/78/19/5538.full#ref-list-1>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://cancerres.aacrjournals.org/content/78/19/5538>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.