RESEARCH

Open Access

Gene regulatory networks reveal sex difference in lung adenocarcinoma

Enakshi Saha[1], Marouen Ben Guebila[1], Viola Fanfani[1], Jonas Fischer[1], Katherine H. Shutta[1,2], Panagiotis Mandros[1], Dawn L. DeMeo[2,3], John Quackenbush[1,2,4] and Camila M. Lopes-Ramos[1,2,3*]

Abstract

Background Lung adenocarcinoma (LUAD) has been observed to have significant sex differences in incidence, prognosis, and response to therapy. However, the molecular mechanisms responsible for these disparities have not been investigated extensively.

Methods Sample-specific gene regulatory network methods were used to analyze RNA sequencing data from non-cancerous human lung samples from The Genotype Tissue Expression Project (GTEx) and lung adenocarcinoma primary tumor samples from The Cancer Genome Atlas (TCGA); results were validated on independent data.

Results We found that genes associated with key biological pathways including cell proliferation, immune response and drug metabolism are differentially regulated between males and females in both healthy lung tissue and tumor, and that these regulatory differences are further perturbed by tobacco smoking. We also discovered significant sex bias in transcription factor targeting patterns of clinically actionable oncogenes and tumor suppressor genes, including *AKT2* and *KRAS*. Using differentially regulated genes between healthy and tumor samples in conjunction with a drug repurposing tool, we identified several small-molecule drugs that might have sex-biased efficacy as cancer therapeutics and further validated this observation using an independent cell line database.

Conclusions These findings underscore the importance of including sex as a biological variable and considering gene regulatory processes in developing strategies for disease prevention and management.

Highlights

- Genes associated with cell proliferation, immune response, and drug metabolism are differentially targeted by transcription factors between males and females in healthy lung tissue and lung adenocarcinoma samples.
- Gene regulatory differences between males and females are also affected by an individual's history of smoking.
- Significant sex bias exists in transcription factor targeting patterns of several clinically actionable oncogenes and tumor suppressor genes including *AKT2* and *KRAS*.
- Analyzing sex-biased gene regulatory networks in lung adenocarcinoma identified small-molecule cancer therapeutics with potential sex-specific efficacy, underscoring the importance of considering sex-specific gene regulation in precision medicine.

*Correspondence:
Camila M. Lopes-Ramos
nhclr@channing.harvard.edu
Full list of author information is available at the end of the article

Saha *et al. Biology of Sex Differences*   (2024) 15:62

Page 2 of 19

Plain English Summary

Lung adenocarcinoma (LUAD) is a disease that affects males and females differently. Biological sex not only influences chances of developing the disease, but also how the disease progresses and how effective various therapies may be. We analyzed sex-specific gene regulatory networks consisting of transcription factors and the genes they regulate in both healthy lung tissue and in LUAD and identified sex-biased differences. We found that genes associated with cell proliferation, immune response, and drug metabolism are differentially targeted by transcription factors between males and females. We also found that several genes that are drug targets in LUAD, are also regulated differently between males and females. Importantly, these differences are also influenced by an individual's smoking history. Extending our analysis using a drug repurposing tool, we found candidate drugs with evidence that they might work better for one sex or the other. These results demonstrate that considering the differences in gene regulation between males and females will be essential if we are to develop precision medicine strategies for preventing and treating LUAD.

Graphical Abstract
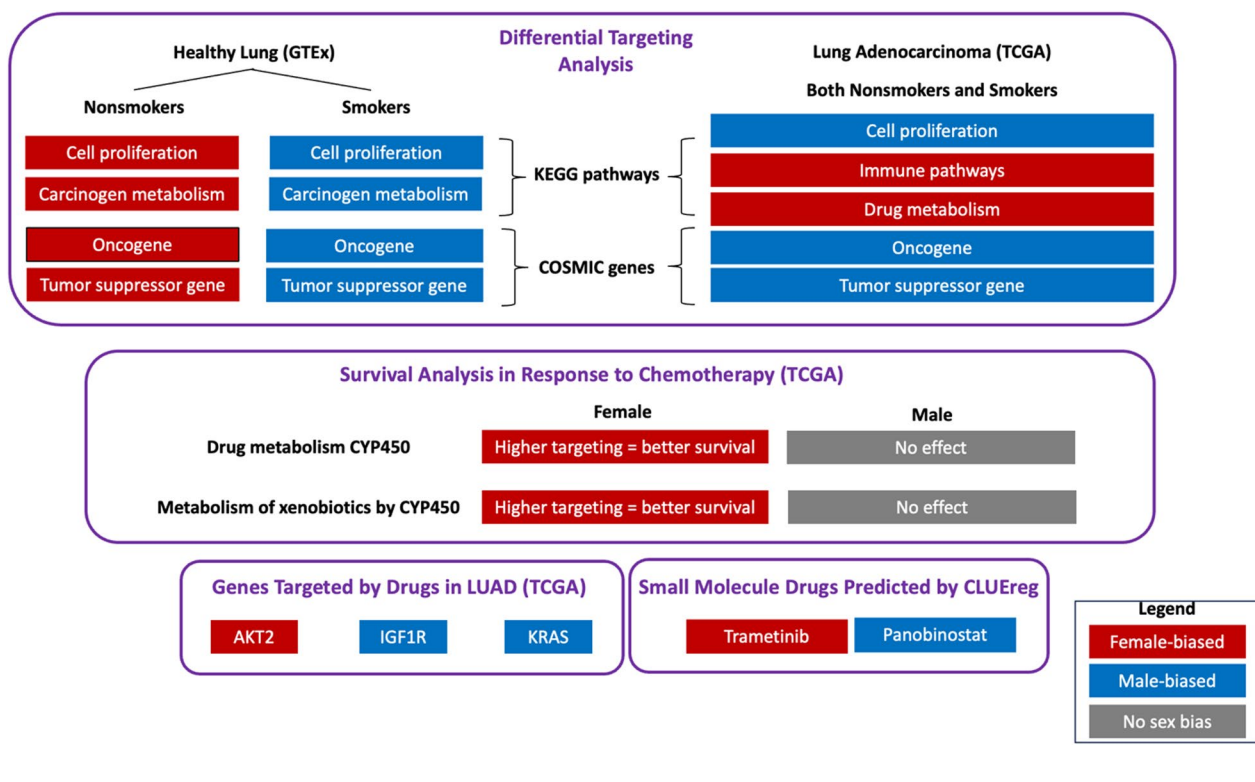
Introduction

Lung adenocarcinoma (LUAD) exhibits significant sex differences in incidence, prognosis, and response to therapy. LUAD has been observed to be more prevalent in females than males [1–3], with the sex difference being more pronounced among nonsmokers (individuals who have never smoked in their lifetime) [4]. However, males with LUAD on average have more severe disease and poorer survival outcomes compared to females with the disease [5]. Treatment responses and toxicity are also influenced by sex [5]; while females usually respond better to chemotherapy than do males [6], immune checkpoint inhibitors have been found to be more effective in males [7] with lung cancer.

Saha *et al. Biology of Sex Differences*   (2024) 15:62

Page 3 of 19

Increased LUAD susceptibility in females may partially be attributed to the effect of estrogens on lung carcinogen metabolism. For example, polymorphisms in cytochrome P450 1A1 (*CYP1A1*) and glutathione *S*-transferase M1 (*GSTM1*) may contribute to the increased risk of females for lung cancer. Females with the *CYP1A1* mutant/*GSTM1* null genotypes face an elevated risk, regardless of their smoking history, potentially influenced by estrogen exposure [8]. Hormonal influences could contribute not only to lung cancer incidence, but also its development and survival outcomes [9]. Prior research has detected the existence of estrogen receptors in malignant lung tissues in both sexes [10]. However, the effects of sex steroid hormones may not account for all differences in how males and females respond to environmental carcinogens, including smoking [4]. Among other factors, higher DNA adduct levels and more frequent mutations in the proto-oncogene *KRAS* in females have also been cited as a possible contributor governing higher lung cancer risk in females [11]. Genetic and metabolic factors have also been cited as potential mediators for the better prognostic outcomes in females [12, 13] compared to males with lung cancer. While previous studies have focused on molecular alterations and gene expression alone [14, 15], an integrative analysis of multi-omics data from a systems perspective can offer valuable insights into sex-specific regulatory mechanisms linked to both lung cancer incidence and clinical outcome.

Despite documented sex differences in LUAD risk and subsequent disease outcome, most methods used in the development and selection of cancer therapeutics do not consider biological sex differences, in part because their molecular drivers are poorly understood, and partly because clinical trials are not designed to address sex-specific effects. Understanding the regulatory processes that differentiate between the sexes in both healthy lung tissue and in LUAD will not only help to elucidate disease mechanisms but also identify more effective therapeutic approaches for both sexes.

We inferred gene regulatory networks using PANDA [16] and LIONESS [17], methods that in combination integrate genome-wide transcription factor binding site maps, transcription factor protein–protein interaction data, and gene expression profiles to produce sample-specific regulatory network models Fig. 1 and which have successfully uncovered sex-specific regulatory drivers of health and disease in previous studies [18–22]. We compared these sample-specific regulatory networks between males and females to identify genes and biological pathways targeted by transcription factors in a sex-biased manner in both healthy lung tissue and in LUAD samples. We further explored how this sex bias is influenced by smoking behavior, a significant risk factor for lung cancer.

As a primary measure of regulatory network differences, we used differential gene targeting, which identifies significant changes in the network model transcription factor repertoire controlling each gene. Among healthy samples, genes associated with cell adhesion and cell proliferation were highly targeted among female non-smokers, while in tumor samples these genes showed higher targeting in males, irrespective of smoking history. Genes associated with immune pathways exhibited higher targeting in tumor samples from females than in those from males, suggesting the potential for sex-based differential response to cancer immunotherapy. Pathways with known relevance in chemotherapy response such as drug metabolism cytochrome P450 (CYP450) showed higher targeting in females, compared to males. Furthermore, an elevated targeting of drug metabolism CYP450 was also associated with favorable survival outcomes in response to chemotherapy among females but not males. We also uncovered significant sex bias in transcription factor targeting of oncogenes and tumor suppressor genes, including *AKT2* and *KRAS* that suggests lung cancer drugs targeting these genes might exhibit differences between the sexes in both efficacy and toxicity. Using an *in-silico* drug repurposing tool, we identified several small-molecule drugs that might have sex-biased efficacy as cancer therapeutics and further validated this hypothesis using an independent cell line database.

Results

Understanding sex difference in incidence risk of LUAD through differential gene regulation in healthy lung tissue

To understand why females have a higher risk of developing LUAD compared to males, especially among non-smokers, we compared male and female gene regulatory networks inferred from GTEx for healthy lung samples (Fig. 2). We identified several key pathways that are targeted by transcription factors in a sex-biased manner in healthy lung that shed light on potential mechanisms driving sex difference in disease risk.

Among nonsmokers from GTEx, biological pathways responsible for cell proliferation, cell adhesion, and cell migration were observed to be highly targeted in females compared to males (FDR < 0.05). Such pathways include the hedgehog signaling pathway, WNT signaling pathway, notch signaling pathway, ERBB signaling pathway, non-small cell lung cancer, focal adhesion and adherens junction (Fig. 2). We validated our findings in healthy lung samples from an independent dataset (LGRC) (Figure D.1), where all these pathways except the hedgehog signaling pathway and the pathway associated with
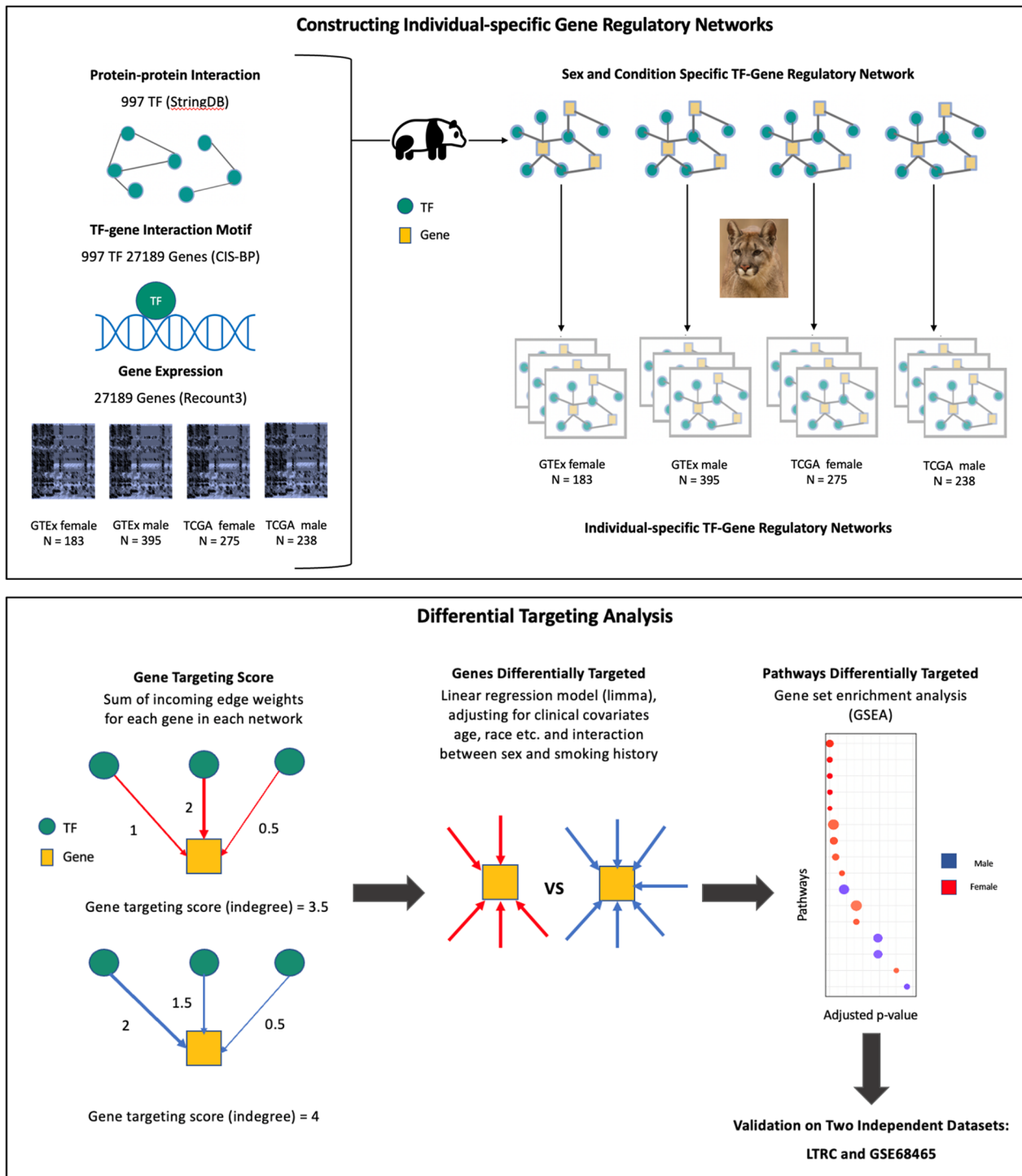
Fig. 1 Schematic overview of the study. Top box, overview of the approach used to construct individual specific gene regulatory networks with PANDA and LIONESS by integrating information on protein–protein interaction between transcription factors (TFs), TF-gene motif binding, and gene expression data of GTEx healthy lung tissues and TCGA lung adenocarcinoma (LUAD) primary tumor samples from Recount3. Bottom box, overview of the differential targeting analysis and independent datasets for validation

Saha *et al. Biology of Sex Differences*   (2024) 15:62

Page 5 of 19
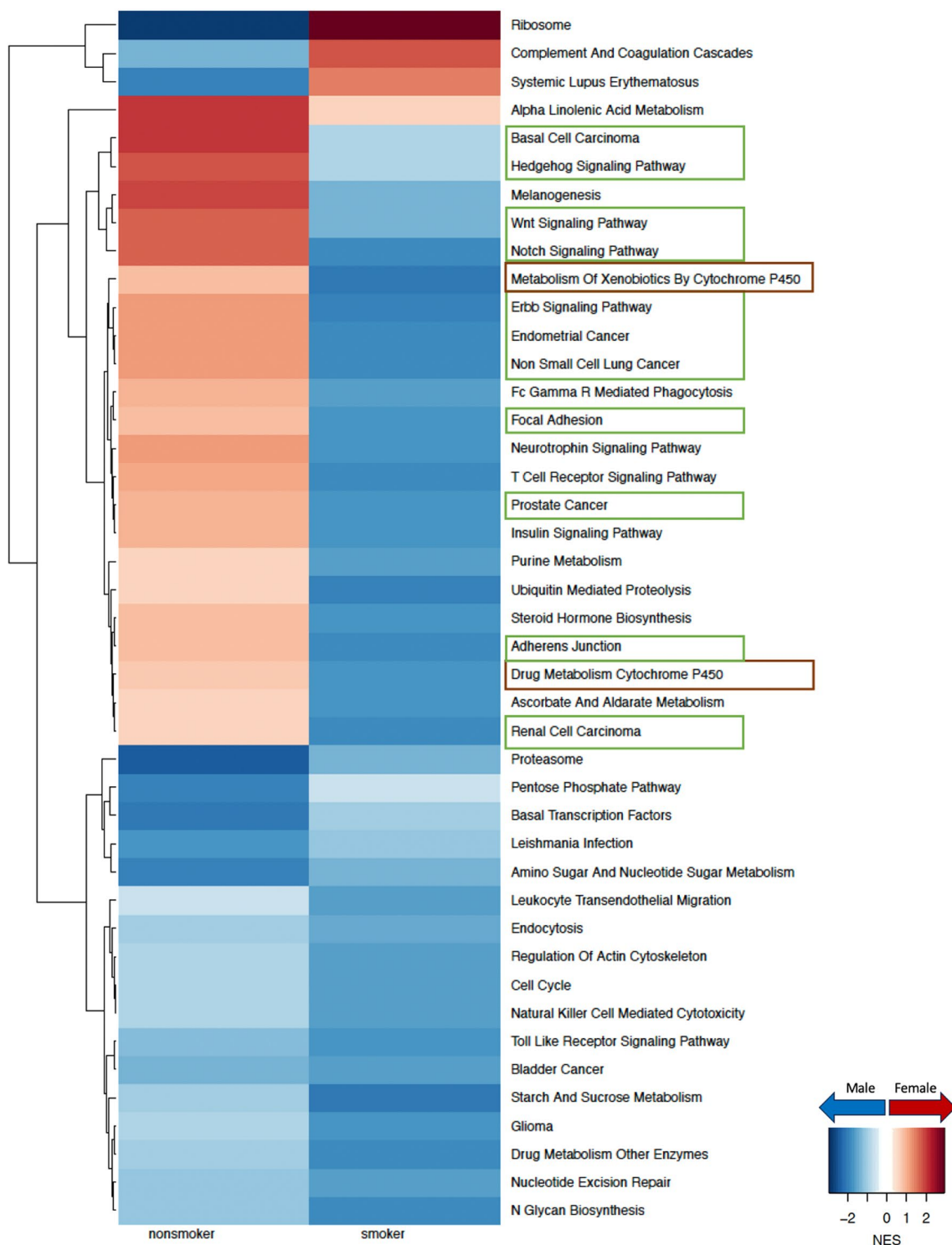
Fig. 2 Sex difference in GTEx healthy lung samples within nonsmokers and smokers. Normalized enrichment scores (NES) from gene set enrichment analysis (GSEA) using KEGG pathways are shown for all pathways that have significant (adjusted p-value < 0.05) sex difference among either nonsmokers or smokers. Pathways with higher targeting in male are marked blue and pathways with higher targeting in female are marked red. Green boxes highlight pathways associated with cell proliferation and brown boxes highlight pathways associated with environmental carcinogen metabolism

Saha *et al. Biology of Sex Differences*   (2024) 15:62

Page 6 of 19

non-small cell lung cancer showed higher targeting among females than males, consistent with the evidence from GTEx.

In contrast, within GTEx smokers, all pathways associated with cell proliferation and cell adhesion mentioned above were more highly targeted in males than females. In the LGRC dataset only two pathways were validated: non-small cell lung cancer and hedgehog signaling pathway (Figure D.1). The CYP450 drug metabolism pathway, which is associated with environmental carcinogen metabolism [11] also had higher targeting in female among nonsmokers and in male among smokers, within both GTEx (Fig. 2) and LGRC (Figure D.1) control samples. Further, our analysis in healthy human lung indicates that pathways related to cell proliferation and environmental carcinogen metabolism are differentially regulated between males and females, which might contribute to the difference in risk of developing LUAD between the sexes.

Understanding sex difference in LUAD prognosis through differential gene regulation

To understand why males have poorer prognosis than females with LUAD, we compared the gene regulatory networks of primary tumors from males and females from the TCGA and identified key pathways with sex-biased targeting by transcription factors. Specifically, we found that pathways involved in cell adhesion, cell proliferation, and cell migration, such as WNT signaling pathway, pathways in cancer, tight junction, and adherens junction, all have higher targeting in tumors from males compared to those from females irrespective of smoking status. It is interesting to note that for nonsmokers (Fig. 3), cell proliferation and migration-related pathways switched from having higher targeting in healthy females to having higher targeting in male tumors. And for smokers (Fig. 4), pathways related to cell proliferation and cell migration that were already highly targeted in healthy males become even more highly targeted in male tumors, compared to females. We replicated this network analysis using an independent LUAD dataset (GSE68465) (Figure D.2) and validated that among nonsmokers, WNT signaling pathway and tight junction were more highly targeted in male tumors than in those from females. We also validated that among smokers, pathways in cancer and adherens junction showed higher targeting among male tumors, consistent with the results from TCGA.

For each pathway we can get the leading genes with most sex differential targeting, and as an example in Figure D.3 we demonstrate the directionality of sex differences in targeting of such leading genes in the ribosomal pathway, which had a strong sex difference that varied by tissue. An interesting point to note is that the

sex bias in gene regulation of biological pathways might vary by the racial background of the population being studied. Since individuals in the TCGA data are mostly of White and African American descent, our findings mentioned in the previous paragraph might not be generalizable to individuals of other races. To demonstrate that we performed a similar analysis on an independent dataset (supplementary material, section F and Figure D.4) consisting of East Asian individuals and found significant disparity in the directionality of sex differences in several key pathways involved in cell proliferation and immune response.

We then turned our attention to oncogenes and tumor suppressor genes cataloged in the COSMIC database [23] and found these to also be highly differentially targeted between the sexes in both healthy and tumor samples (Fig. 5). Among nonsmokers in healthy GTEx lung samples, both oncogenes and tumor suppressor genes showed higher targeting (p-value of Wilcoxon signed rank test is 2.229e−09 for oncogenes and 3.614e−05 for tumor suppressor genes) in females compared to males. Whereas among the nonsmokers in the TCGA tumor samples, both oncogenes and tumor suppressor genes showed higher targeting in male samples (p-value of Wilcoxon signed rank test is 2.334e−09 for oncogenes and 5.217e−07 for tumor suppressor genes), which may help explain poorer prognosis in males compared to females. For smokers, oncogenes and tumor suppressor genes showed higher targeting for males than females in both healthy lung samples from GTEx (p-value of Wilcoxon signed rank test is 3.546e−08 for oncogenes and 2.296e−12 for tumor suppressor genes), as well as LUAD tumors from TCGA (p-value is 5.906e−08 for oncogenes and 2.296e−12 for tumor suppressor genes).

To understand whether sex differences in regulation of biological pathways might explain poorer survival among males with LUAD, we performed survival analysis on TCGA data using a Cox proportional hazard model for each of these pathways. We found a higher targeting of the RNA Degradation pathway to be associated with poorer survival outcome in males (z-score of the coefficient corresponding to pathway score is 2.030 with p-value 0.042) but did not have any impact in females (z-score of the coefficient corresponding to pathway score is -0.740 with p-value 0.459). The leading genes contributing towards a higher targeting of RNA degradation among males include *CNOT1* [24], *CNOT2* [25], *CNOT3* [26] and *DCP1A* [27], all of which have previously been found to have prognostic significance in various cancers, including non-small cell lung cancer.

Fig. 3 Sex difference among nonsmokers in GTEx healthy lung and in TCGA LUAD. Normalized enrichment scores (NES) from GSEA using KEGG pathways are shown for all pathways that have significant (adjusted p-value < 0.05) sex difference among either TCGA nonsmokers or TCGA smokers. Pathways with higher targeting in male are marked blue and pathways with higher targeting in female are marked red. Green boxes highlight pathways associated with cell proliferation and purple boxes highlight pathways associated with immune response
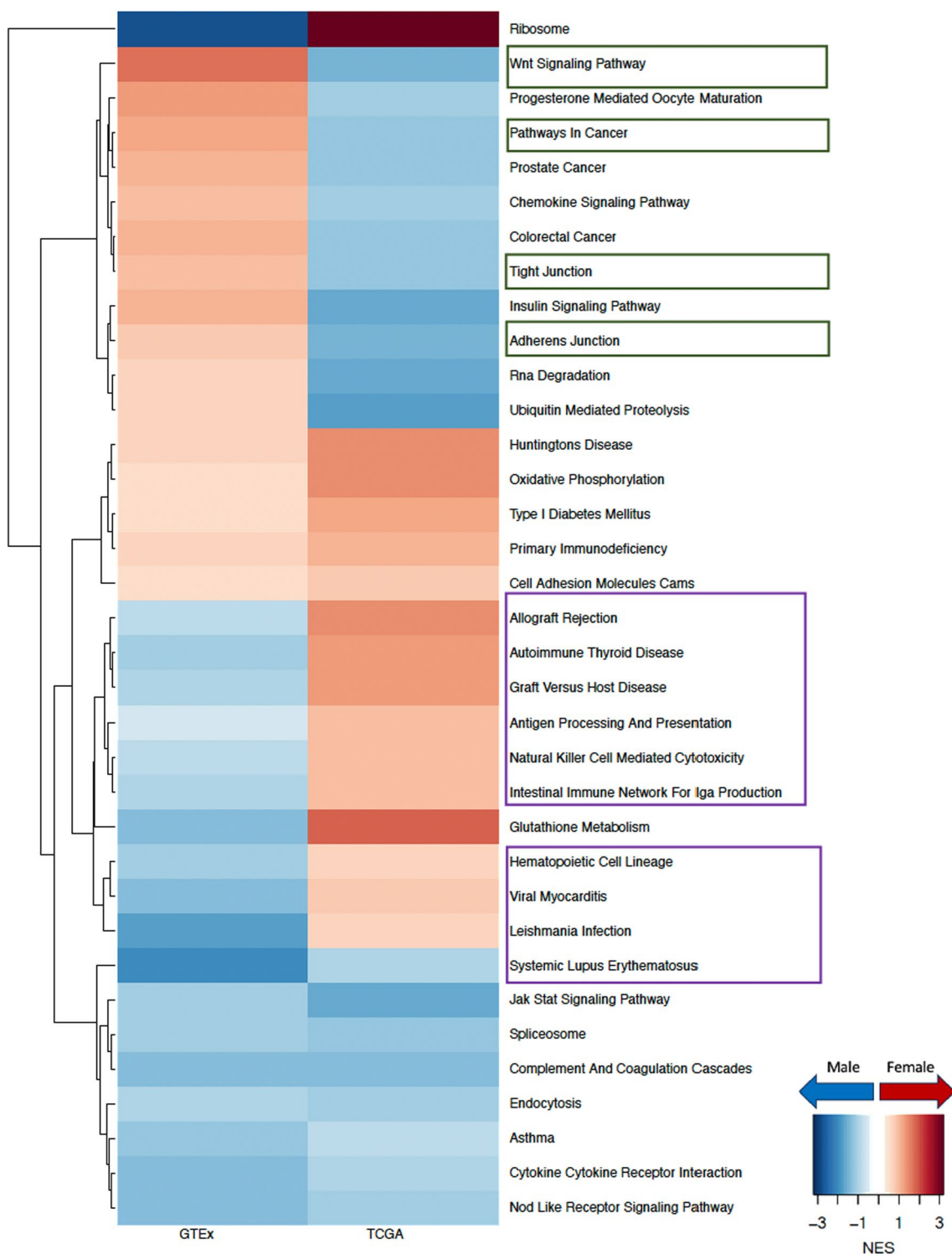
Fig. 4 Sex difference among smokers in GTEx healthy lung and in TCGA LUAD. Normalized enrichment scores (NES) from GSEA using KEGG pathways are shown for all pathways that have significant (adjusted p-value < 0.05) sex difference among either TCGA nonsmokers or TCGA smokers. Pathways with higher targeting in male are marked blue and pathways with higher targeting in female are marked red. Green boxes highlight pathways associated with cell proliferation and purple boxes highlight pathways associated with immune response
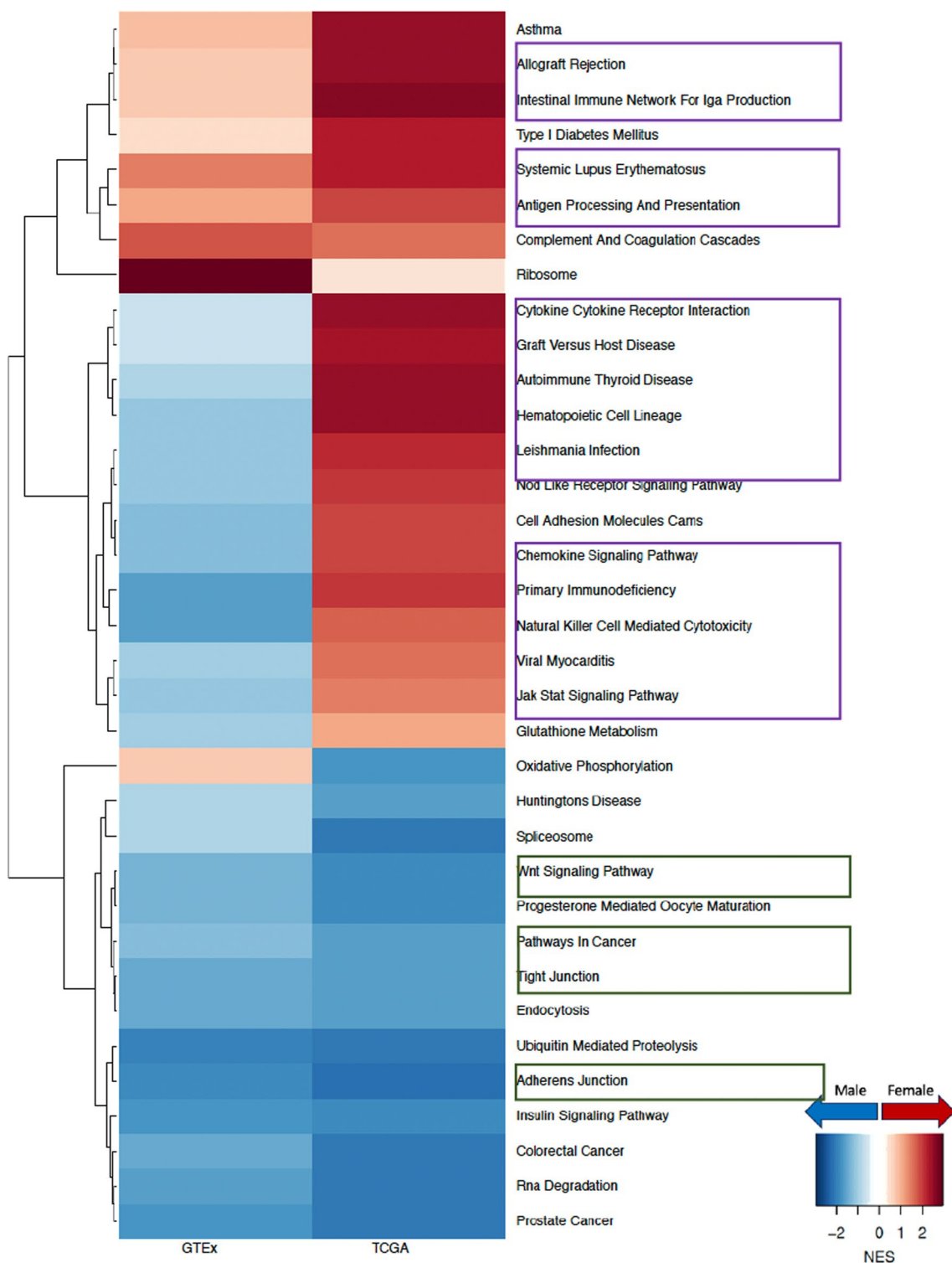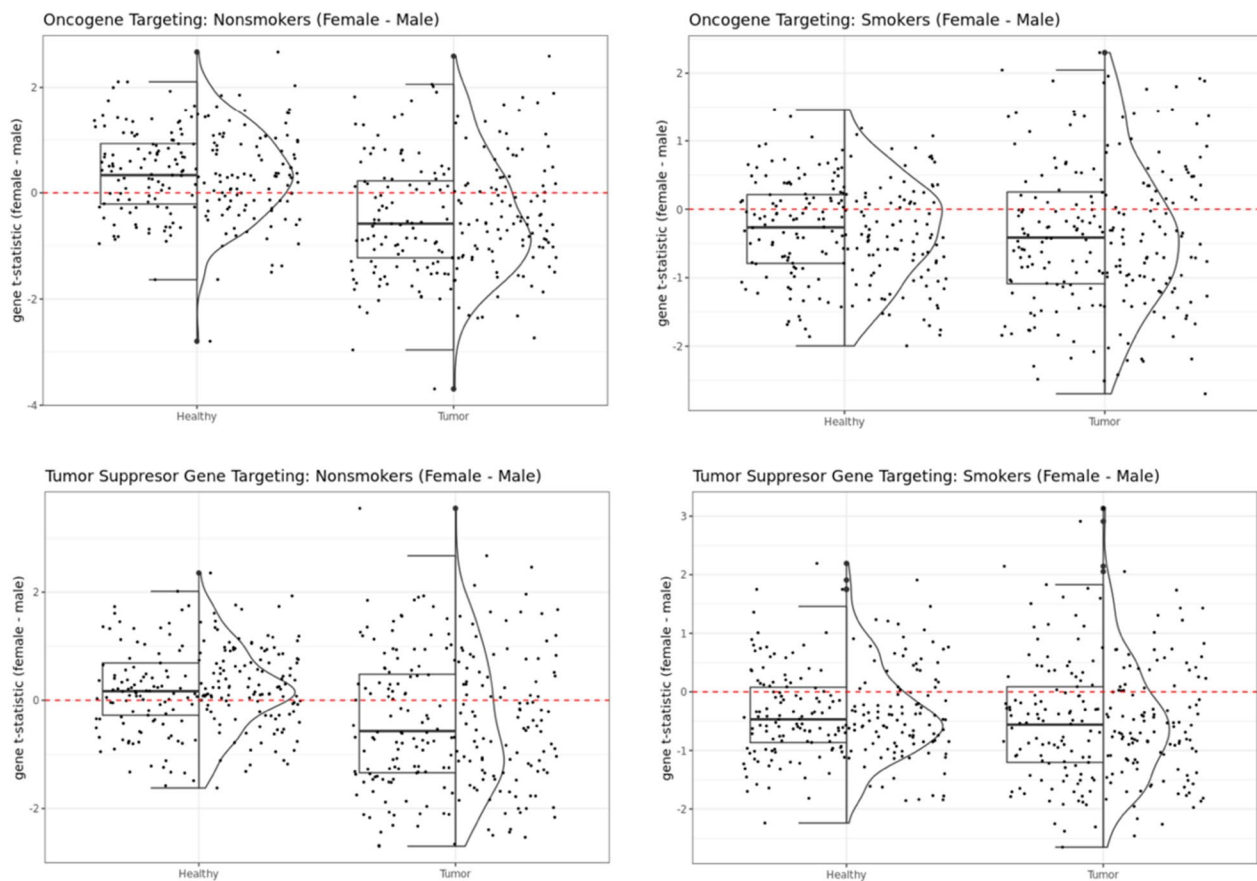
Saha *et al. Biology of Sex Differences*   (2024) 15:62

Page 9 of 19

Fig. 5 Sex difference in targeting of oncogenes (top row) and tumor suppressor genes (bottom row) in GTEx and TCGA nonsmokers (left column) and smokers (right column)

Sex difference in immunotherapy

The TF-targeting of almost all immune pathways is higher in tumor samples from females than those from males. This female-bias is particularly pronounced among TCGA tumor samples from smokers, where we observed that immune-related pathways including allograft rejection, intestinal immune response for IGA production, systemic lupus erythematosus, and antigen processing and presentation, all showed considerably higher targeting in females (Fig. 4).

Although among GTEx nonsmokers (Fig. 5) these pathways were more highly targeted in males, in TCGA nonsmokers we found a shift towards higher targeting in females—except for systemic lupus erythematosus which remained highly targeted in male tumor samples. Additionally, other immune pathways such as hematopoietic cell lineage and natural killer cell mediated cytotoxicity initially showed higher targeting in GTEx males, but switched to higher targeting in TCGA females, irrespective of smoking status. This female-bias in targeting of immune pathways was further validated in tumor samples from smokers in GSE68465.

To understand whether sex-difference in TF-targeting of immune pathways can be partially attributed to a sex-difference in immune cell infiltration, we performed immune cell type deconvolution analysis of TCGA data. We found that, consistent with a higher targeting of immune pathways in females, various immune cell proportions including natural killer cells, CD4+naive T cells, myeloid dendritic cells, and B cells were more highly targeted in female tumor samples than male tumor samples (Fig. 6). The only exceptions were CD4+Th2 helper cells that are present in higher proportions among male samples. Differential targeting of immune pathways, along with a sex-biased infiltration of immune cells, might contribute to varying degrees of efficacy of immune checkpoint inhibitors shown to exist among males and females with LUAD (Table C.1) [28]. However, in healthy GTEx samples, we did not find any sex difference in the proportion of immune cells that had sex-biased infiltration rate in TCGA (Figure D.5) except for natural killer T cells, which showed higher proportion in males compared to females among nonsmokers.
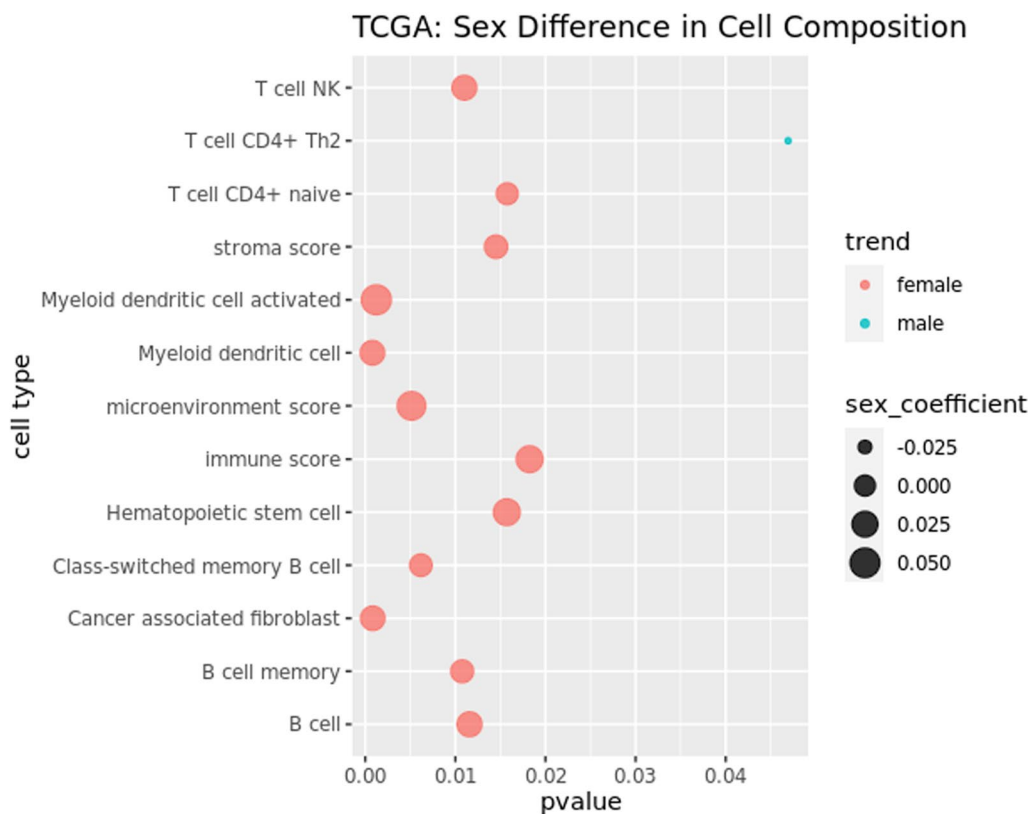
Fig. 6 Sex Difference in immune and stromal cell composition in TCGA LUAD samples. Cell compositions are computed using "xcell", which derives cell composition proportion of 36 immune and stromal, along with three composite scores: immune score, stroma score and microenvironment score. The bubbleplot shows only those cells that are significantly (p-value < 0.05) different in proportion in male and female tumor samples

Sex difference in chemotherapy

There is empirical evidence of significant sex differences in chemotherapy response [29] in LUAD, with females having better outcomes than males in most cases [6]. To explore this, we used networks only for those patients who received chemotherapy and fit a Cox proportional hazard model to identify pathways with a sex-biased impact on survival.

Within females, greater targeting of two CYP450 pathways—drug metabolism (p-value 0.016) and metabolism of xenobiotics (p-value 0.052)—was associated with better survival, while in males a differential targeting of these pathways did not have any impact on survival (p-value for metabolism of xenobiotics by CYP450 was 0.110 and p-value for drug metabolism CYP450 was 0.157). This same pattern of influence on the interaction between drug metabolism CYP450 targeting, and chemotherapy treatment has been reported in colon cancer [20]. Notably, these pathways did not have any significant impact on survival in treatment-naïve tumor samples, which indicates that gene regulatory network analysis has the power to predict the potential for individuals to respond to

clinical interventions, including the use of chemotherapy agents.

Sex difference in targeted therapy

Cancer therapeutics targeting specific genes have also been observed to have a sex-biased impact on both dose-efficacy and dose-toxicity [30]. To understand how differential regulation of specific drug targets might contribute towards different efficacy of various cancer drugs in males and females with LUAD, we chose 28 genes commonly targeted by lung cancer drugs [31] for an in-depth analysis (Fig. 7). Among these genes, three showed significant (p-value less than 0.05) sex-bias in transcription factor targeting patterns: within nonsmokers *AKT2* showed higher targeting among females; *KRAS* and *IGF1R* showed higher targeting among males compared to females, irrespective of smoking status.

We then used CLUEreg [32], a tool designed to match disease states to potentially therapeutic small molecule drugs based on differential regulation between tumor and healthy samples, to identify potential targeted cancer therapeutics that might be more beneficial to individuals of one sex over the other, and derived a list of small
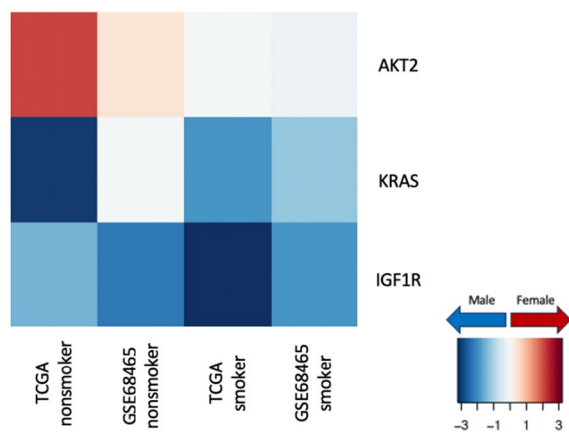
Fig. 7 Sex difference in transcription factor targeting of genes commonly targeted by drugs in lung cancer in TCGA and validation data GSE68465, split by smoking status. The heatmap shows t-statistics corresponding to the sex coefficient from a limma analysis on the gene targeting score (indegree) (p-value < 0.05 for the sex coefficient). Genes with higher targeting in male samples are marked in blue and genes with higher targeting in female samples are marked in red

molecule drug candidates for both males and females (Full list of drugs for males and females are available as supplementary materials S5 and S6). After cross-referencing these candidate drugs with the Genomics of Drug Sensitivity in Cancer (GDSC) database, we identified several small molecule drugs that might be beneficial for either males or females with LUAD. While several conventional cancer therapeutics such as Tanespimycin and Cisplatin appeared as potential drug candidates for both sexes, we found three drug candidates (Trametinib, Scriptaid/Vorinostat and Actinomycin-d/Dactinomycin) that had evidence of potential efficacy exclusively for females and one drug candidate (LBH-589/Panobinostat) exclusively for males; all four of these drugs are FDA approved. All four drugs except Dactinomycin (which had a tau value of 0.0632) had a tau value of 0.0006 or lower, as calculated by CLUEreg, which suggests that these drugs have specific, rather than pleiotropic effects as compared to the other drugs in the database.

Using the GDSC dataset, we validated that female cell lines had greater sensitivity for Trametinib (p-value 0.00027 Mann–Whitney test), and male cell lines were more sensitive to Panobinostat (p-value 0.01396 Mann–Whitney test), as predicted by CLUEreg (Fig. 8). We did not, however, find supporting evidence for sex differences in the efficacy of Vorinostat or Dactinomycin. This may be due to the relatively small number of cell lines of either sex that have been profiled and the innate variability among individuals in regulatory potential. Although preliminary, the validation of CLUEreg drug predictions using an independent cell line drug screening dataset
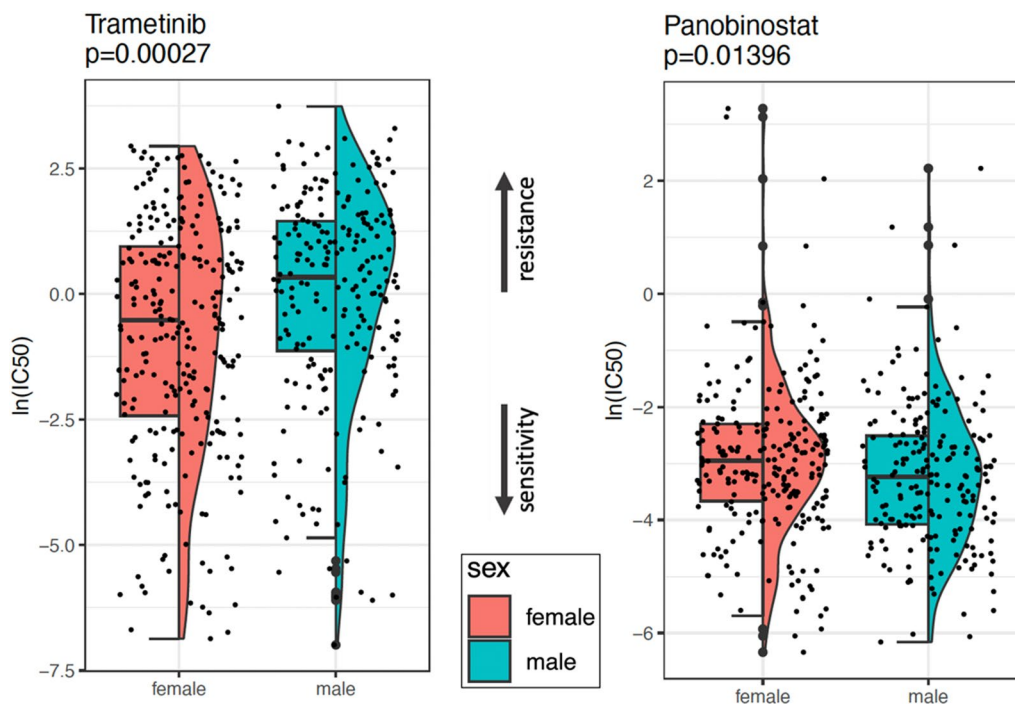

Fig. 8 Validation of sex-specific therapeutics predicted by CLUEreg using GDSC drug sensitivity data. Boxplots of half maximal inhibitory concentration values (Log IC50) for male and female cell lines treated with Trametinib and Panobinostat, Mann-Whitney test

Saha *et al. Biology of Sex Differences*   (2024) 15:62

Page 12 of 19

Discussion

LUAD, like many cancers, is known to differ between males and females in disease risk, development, progression, and response to therapy. While lifestyle differences, androgen and estrogen levels, and the genetic effects of different allosomes may play some role, the causes of these apparent sex differences remain largely unstudied. Although there are some differences in gene expression between males and females, both in healthy and tumor samples, these are largely confined to the sex chromosomes [20] and do not shed much light on mechanistic differences that might help explain the well-established clinical differences.

Despite the lack of differential expression, our hypothesis was the reported differences in LUAD between the sexes was reflected in sex-biased patterns of gene regulation. We inferred gene regulatory networks using PANDA and LIONESS and compared the networks to identify sex-specific regulatory patterns in healthy and LUAD samples that might provide mechanistic explanations for sex-specific phenotypic differences. Using differential targeting analysis on individual-specific gene regulatory networks, we identified sex-bias in transcription factor targeting of biological pathways associated with cell proliferation, environmental carcinogen metabolism and immune response in healthy lungs, as well as in LUAD.

We found differences in regulatory processes controlling genes involved in cell proliferation and adhesion pathways, including many implicated in cancer, such as the hedgehog signaling pathway [33], WNT signaling pathway [34], notch signaling pathway [35] and ERBB signaling pathway [36]. Within healthy samples these pathways showed higher targeting in female nonsmokers and male smokers, whereas within tumor samples all these pathways were highly targeted in males, irrespective of smoking status. These differences in gene regulation may explain why females have a greater risk of developing LUAD, but the disease trajectory in males leads to more rapid progression and poorer outcomes.

Chemotherapy drugs such as carboplatin and paclitaxel has been observed to have sex differences in both efficacy and toxicity in non-small cell lung cancer, where females have a more favorable prognosis than males [6]. Our analysis suggests that the differential response to chemotherapeutic agents might be associated with differential targeting of drug metabolism CYP450 pathways in gene regulatory networks. Among patients undergoing chemotherapy, we found that higher targeting of two CYP450 pathways, namely drug metabolism and xenobiotics metabolism, was associated with improved survival in females, while in males, differential targeting of these pathways did not have any significant impact on survival. A similar influence of drug metabolism CYP450 targeting on chemotherapy outcomes was previously reported in the context of colon cancer [20].

Not surprisingly, we also found sex-specific differences in the regulation of immune related processes, as well as proportion of infiltration of various immune cells within tumor samples. Not only do this shed light on cancer prognosis but might also elucidate towards a sex-biased response to various cancer immunotherapies [28], including PD1 and PDL1 inhibitors.

We also identified that several genes as differentially targeted between the sexes and for which directed therapies exist, including *AKT2, IGF1R,* and *KRAS*. While these genes have been extensively studied, there are virtually no published studies on potential sex differences in response to drugs targeting these genes. However, it has been shown in a murine model that drugs targeting *IGF1R* (Insulin-like Growth Factor-1) improve lifespan with a reduction of neoplasm only in females [37], which aligns with our findings.

We identified four FDA-approved small-molecule drug candidates that might have a sex-biased efficacy: three drugs (Trametinib, Vorinostat and Dactinomycin) were identified exclusively for females and Panobinostat was identified exclusively for males. Using an independent database, we validated that female cell lines had indeed higher sensitivity for Trametinib, and male cell lines had higher sensitivity for Panobinostat. Trametinib targets *MAP2K1* [38], which showed higher targeting in males than females, based on our analysis of regulatory networks. Higher targeting of *MAP2K1* by transcription factors may reduce the effectiveness of cancer therapeutics targeting *MAP2K1* such as trametinib in males compared to females. Panobinostat is a histone deacetylase (HDAC) inhibitor [39]. HDAC inhibitors cause upregulation of the cell cycle gene *CDKN1A*, leading to cell cycle arrest [40, 41]. *CDKN1A* showed higher targeting by transcription factors in females than males. Higher targeting of *CDKN1A* by transcription factors may reduce the effectiveness of HDAC inhibitors such as Panobinostat in females compared to males. Although we could not verify the validity of the predictions from CLUEreg on human trials since clinical trials in general do not report drug efficacy by sex, the validation of CLUEreg drug predictions using an independent cell line drug screening dataset underscores the potential of using gene regulatory networks to identify sex-specific cancer therapeutics.

It must be acknowledged that although we adjusted for various clinical confounders such as age, race, smoking

history, and clinical tumor stage in our differential targeting analysis, outcomes might still be influenced by other factors including cellular and genetic heterogeneity, or unobserved clinical phenotypes and risk factors including the effect of hormones, lifestyle habits, environmental exposures, and family history. To establish causal conclusions regarding the effect of regulatory sex-differences in disease mechanism, further work would be required to elucidate the relative contributions, as well as possible interactions between these factors and sex-biased gene regulatory patterns identified by our analysis. Although our analysis suggests a possible impact of cigarette smoking on the sex bias in gene regulatory patterns, the extent to which smoking-related effects may confound or interact with sex-specific differences in gene expression also needs further exploration that would include accounting for the distinctions between former and current smokers, duration of smoking cessation for former smokers, and the number of pack years smoked.

Another limitation of our work is that the discovery datasets from GTEx and TCGA, as well as the validation datasets considered in our analysis, predominantly consist of white and African American individuals. Even though our analyses were adjusted for the impact of race, the generalizability of our findings to other ethnicities might still be limited due to a lack of representation. Indeed, as demonstrated in the supplementary material, section F, we found that the results from TCGA differ from those we found analyzing an independent cohort consisting only of East Asian individuals. In this regard, future studies should be expanded to include more diverse populations if we are to ensure the validity of genomic findings in all individuals.

As we continue to develop methods for inferring gene regulatory, we hope to extend the analyses presented here by exploring how interactions between transcription factor binding, post-translational modifications, and differences in protein activity, as well as epigenetic changes, might alter disease trajectories in a sex-specific manner. Recently, interesting sex differences in protein signaling networks were found in LUAD tumor samples from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) [42] and such data provide a complementary means of exploring drivers of sex differences to the work we are continuing to perform.

Perspectives and significance

Our study highlights the substantial sex differences in gene regulatory patterns in healthy lung and lung adenocarcinoma, as well as how smoking affects gene regulation in males and females. The regulatory differences not only help to explain sex-biases in disease susceptibility and prognosis, but also hold promise for shaping sex-specific therapeutic strategies with the potential to improve outcomes. This underscores the value of using sex-specific alterations in regulatory networks to adapt disease treatment based on each patient's sex as a cornerstone of precision medicine in LUAD as well as other diseases.

Method

Discovery dataset

We downloaded uniformly processed RNA-Seq data from the Recount3 database [43] for two discovery datasets using the R package "recount3" (version 1.4.0) on May 26, 2022: (i) healthy lung tissue samples from the Genotype Tissue Expression (GTEx) Project [44] (version 8) and (ii) lung adenocarcinoma (LUAD) samples from The Cancer Genome Atlas (TCGA) [45]. Clinical data for GTEx samples were accessed from the dbGap website (https://dbgap.ncbi.nlm.nih.gov/) under study accession phs000424.v8.p2. Clinical data for TCGA samples were downloaded from Recount3. Throughout our analysis the GTEx samples will be referred to as "healthy lung samples."

From 655 healthy lung samples in GTEx, we removed 77 samples because they were designated as "biological outliers" in the GTEx portal (https://gtexportal.org/) for various reasons (as described in https://gtexportal.org/home/faq). The remaining 578 samples (395 males, 183 females) were used in the analysis. We verified that the self-reported gender for GTEx samples aligned with the biological sex through a principal component analysis (PCA) of gene expression values of 36 genes on the Y chromosome (Figure D.6).

From the TCGA dataset, we removed two recurrent tumor samples and 59 samples from normal adjacent tissues, keeping only primary tumor samples. For individuals with multiple samples, we retained the sample with the highest sequencing depth. Finally, we also removed two samples annotated as "female" as these samples clustered with "male" samples using PCA for the Y chromosome as above (Figure D.6). We also removed one sample with missing gender information. Subsequent analyses were performed on the remaining 513 primary lung adenocarcinoma tumor samples (238 males, 275 females).

We extracted TPM normalized gene expression data from both GTEx and TCGA using the "getTPM" function in the Bioconductor package "recount" (version 1.20.0) [46] in R (version 4.1.2). We excluded lowly expressed genes by removing those with counts < 1 TPM in at least 10% of the samples in GTEx and TCGA combined, thus removing 36,360 annotated genes, and leaving 27,495 (including 36 Y genes and 884 X genes) genes for analysis. To build gene regulatory networks, we kept only genes that were present both in this filtered gene set and,

Saha *et al. Biology of Sex Differences*   (2024) 15:62

Page 14 of 19

Table 1 Clinical characteristics of the discovery and validation datasets

	GTEx (healthy lung)	TCGA (LUAD tumor)	LGRC (healthy lung)	GSE68465 (LUAD tumor)
Sample size	578	513	108	432
Sex				
Female (%)	183 (31.66%)	275 (53.61%)	59 (54.63%)	214 (49.54%)
Male (%)	395 (68.34%)	238 (46.39%)	49 (45.37%)	218 (50.46%)
Age				
Mean ± std (range)	54 ± 11.84 (21–70)	65 ± 10.05 (33–88)	64 ± 11.35 (32–87)	64 ± 10.09 (33–87)
Race				
White (%)	493 (85.29%)	388 (75.63%)	–	289 (66.90%)
Black or African American (%)	70 (12.11%)	50 (9.75%)	–	12 (2.78%)
Others (%)	15 (2.60%)	9 (1.75%)	–	6 (1.39%)
Unknown (%)	-	66 (12.87%)	–	125 (28.93%)
Smoking status				
Smokers (%)	382 (66.09%)	424 (82.65%)	65 (60.19%)	290 67.13%)
Never-smokers (%)	180 (31.14%)	75 (14.62%)	32 (29.63%)	48 (11.11%)
NA (%)	16 (2.77%)	14 (2.73%)	12 (10.18%)	94 (21.76%)
Tumor stage				
I (%)	–	274 (53.41%)	–	148 (33.26%)
II (%)	–	121 (23.59%)	–	242 (56.02%)
III (%)	–	84 (16.37%)	–	28 (6.48%)
IV (%)	–	26 (5.07%)	–	12 (2.78%)
NA (%)	–	8 (1.56%)	–	2 (0.46%)
Ischemic time (hours)				
Mean ± std (range)	8.02 ± 6.98 (0.0–24.4)	–	–	–

Clinical characteristics by sex are recorded in supplementary tables E.1–E.4

in the TF-target gene regulatory prior used in PANDA and LIONESS (see section "Differential targeting analysis using single-sample gene regulatory networks"). The remaining 27,189 genes, including genes on the sex chromosomes, were used for network inference and analysis. For female samples in both GTEx and TCGA, some genes on the Y chromosome have expression values due to mismapping of transcripts; we manually set Y chromosome gene expression values to "NA" for biological females in both data sets.

Validation dataset
We identified two independent studies from the Gene Expression Omnibus (GEO) for use in validating our findings: GSE47460 (hereafter referred to as LGRC) [47] and GSE68465 [48]. From the LGRC (downloaded on Feb 12, 2023) data, we used 108 samples (59 female and 49 male) annotated as "control" samples for validation. Gene expression data came from the Lung Genomics Research Consortium (LGRC) representing a subset of tissue samples from the Lung Tissue Research Consortium (LTRC) that showed no chronic lung disease by CT or pathology. This study used the Agilent-014850 Whole Human Genome Microarray 4×44K G4112F and

Agilent-028004 SurePrint G3 Human GE 8×60K Microarray for gene expression profiling. Data from GSE68465 (downloaded on Jan 24, 2023) consisted of gene expression for lung adenocarcinoma primary tumor samples from 462 individuals. This study used Affymetrix Human Genome U133A Array for gene expression profiling. Nineteen samples were removed because of missing gender information. We also removed six samples annotated as "female" and five samples annotated as "male" based on PCA of expression of 65 Y genes (Figure D.6). The remaining 432 samples (218 male and 214 female) were used in the final validation analysis.

Normalized expression data and clinical data were downloaded using the R package "GEOquery" version 2.62.2. For genes with multiple probe sets, we kept the probe with the highest standard deviation in expression across samples and the gene set was further filtered to remove any genes that did not overlap with those in the TF/target gene regulatory network prior. This left 13,575 genes in GSE47460 (LGRC) and 13,516 genes in GSE68465 that were used in subsequent analyses. The LGRC data did not show any batch effect and so no correction was used. The GSE68465 dataset contained LUAD specimens from the following sources: University

of Michigan Cancer Center (100 samples), University of Minnesota VA/CALGB (77 samples), Moffitt Cancer Center (79 samples), Memorial Sloan-Kettering Cancer Center (104 samples), and Toronto/Dana-Farber Cancer Institute (82 samples). A principal component analysis on the gene expression data demonstrated distinct clusters corresponding to these sample source, thus exhibiting a strong batch effect; expression data was subsequently batch-corrected using the "ComBat" function implemented in the R package "sva" (version 3.42.0).

Table 1 depicts the clinical characteristics of all the discovery and validation datasets and supplementary tables E.1–E.4 present clinical characteristics by sex.

Differential targeting analysis using single-sample gene regulatory networks

We used PANDA [16] and LIONESS [17] to construct gene regulatory networks (Fig. 1) for each sample in the discovery and validation datasets, using Python package netzooPy [49] version 0.9.10. In addition to the gene expression data obtained from the discovery and validation datasets, two other types of data were integrated to construct the networks: TF/target gene regulatory prior (derived by mapping TF motifs from the Catalog of Inferred Sequence Binding Preferences (CIS-BP) [50] to the promoter of their putative target genes) and protein–protein interaction data (using the interaction scores from StringDb v11.5 [51] between all TFs in the regulatory prior). Our TF/target gene regulatory prior consisted of 997 TFs targeting 61,485 ensemble gene IDs, corresponding to 39,618 unique gene symbols (HGNC), and the protein–protein interaction data contained the measure of interactions between these 997 TFs. We used sex-specific binary motif priors (1 representing the presence of a TF motif and 0 representing the absence of a TF motif on the promoter region of the gene) for males and females, where the male and female motifs were the same for autosomal and X chromosome genes, but motifs on the Y chromosome genes were set to 0 in the female prior. The procedure for deriving the motif prior and the PPI priors are given in the supplementary material. Regulatory networks were constructed for each of the discovery datasets and validation datasets separately for female and male samples. The final networks contained only genes overlapping between the TF/target gene motif prior and the corresponding gene expression dataset.

For each sample's gene regulatory network, we computed the targeting score (or, in-degree) for each gene, which corresponds to the sum of incoming edge weights from all TFs to this gene. Gene targeting scores were compared between males and females using a linear regression model, while adjusting for relevant covariates: sex (Male and Female), race (White, Black or African American, Others and Unknown), age, smoking status (Ever-smoker and Never-smoker) and ischemic time for GTEx; sex (Male and Female), race (White, Black or African American, Others and Unknown), age, smoking status (Ever-smoker and Never-smoker) and tumor stage (stages I, II, III, IV and "NA") for TCGA; using the R package limma (version 3.50.3) [52] and accounting for interaction between sex and smoking history (ever-smokers and never-smokers).

In the LGRC dataset we adjusted for age and smoking status and in GSE68465 we adjusted for age, race, tumor stage and smoking status, while simultaneously considering interaction between sex and smoking history (ever-smokers and never-smokers) for each validation dataset.

Although to model gene regulatory networks, we used sex-specific priors to allow for the presence of Y chromosome genes in males but not in females, we did not include genes on the Y chromosome in our differential targeting analyses while performing a direct comparison between the sexes as this was infeasible due to the absence of the Y chromosome in females. However, we included the *XIST* gene because several male samples showed nonzero expression of *XIST*, especially among the tumor samples from TCGA (Figure D.7).

Pathway Enrichment analysis

A gene set enrichment analysis was performed separately for individuals with different smoking histories using the ranked t-statistics of the coefficient for sex derived from the limma analysis (Fig. 8). We used pre-ranked Gene Set Enrichment Analysis (GSEA) in the R package "fgsea" (version 1.20.0) [53] and gene sets from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database [54] ("c2.cp.kegg.v2022.1.Hs.symbols. gmt"), downloaded from the Molecular Signatures Database (MSigDB) (http://www.broadinstitute.org/gsea/msigdb/collections.jsp). Only gene sets of sizes greater than 15 and less than 500 were considered, after filtering out genes which are not present in the expression dataset, which limited our analysis to 176 gene sets. Multiple testing corrections were performed using the Benjamini–Hochberg procedure [55].

Survival analysis

For each biological pathway, the pathway targeting score was computed as the mean indegree of all genes in the pathway. For survival analysis we used the R package "survival" (version 3.2.13) and fit Cox proportional hazard model ("coxph") for the TCGA data to investigate the effect of transcription factor targeting of different KEGG pathways on survival outcome, while adjusting for age, sex, race, smoking status, tumor stage, and chemotherapy status (yes, no and "NA"). Supplementary Table E.5

Saha *et al. Biology of Sex Differences*   (2024) 15:62

Page 16 of 19

shows distribution of clinical variables among individuals who received chemotherapy versus those who did not.

Immune infiltration analysis

We used "xcell" [56] on the TPM-normalized GTEx and TCGA gene expression data with R package "immunedeconv" (version 2.1.0) to infer immune and stromal cell composition in tumor samples. For every cell type, to quantify whether cell type proportion in tumor are variable by sex, we fit a linear model to predict cell type proportion by sex, while adjusting for age, race, smoking status, and clinical tumor stage.

Finding small molecule drugs with CLUEreg

We identified genes that are differentially targeted between tumor and healthy samples, using linear models on gene targeting scores from GTEx and TCGA data through R package "limma". We accounted for the interaction between sex and disease status (tumor versus healthy), while adjusting for clinical covariates that were available for both GTEx and TCGA, including sex, age, race, and smoking status. Genes were ranked by the adjusted p-values (smallest to largest) from the limma analysis and all genes significantly differentially targeted (at FDR cutoff 0.05) were chosen for males and females separately. The selected differentially targeted genes were split between "high" and "low" targeted based on whether they were more highly targeted in tumor (high) samples or in healthy (low) samples and subsequently used as input to CLUEreg [32] (https://grand.networkmedicine.org/), a web application designed to match disease states to potential small molecule therapeutics, based on the characteristics of the regulatory networks. CLUEreg produced a list of 100 small molecule drug candidates most suitable for reversing the gene targeting patterns in tumor to resemble the gene targeting patterns in healthy samples.

To validate CLUEreg predictions, we used gene expression and drug response data from cancer cell lines in the Genomics of Drug Sensitivity in Cancer (GDSC) [57] dataset, removing cell lines from reproductive cancer types. We classified cell lines as male (n = 227) or female (n = 264) groups considering both expression of the Y chromosome genes (gene expression data from GDSC) and the reported gender of the individual from whom the cell line was derived (Sanger Cell Model Passports, https://cellmodelpassports.sanger.ac.uk/downloads). To test whether drug sensitivity varies by sex, we combined technical replicates by median of log IC50 and compared the log IC50 values reported by GDSC (half maximal inhibitory concentration) between male and female cell lines using Wilcoxon-Mann–Whitney test.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13293-024-00634-y.

Supplementary Material 1 A. Designing Sex-specific Transcription Factor-Gene Motif Prior. B. Designing Protein-protein Interaction Prior. C. Sex Difference in anti PD-1 and anti PDL-1 Inhibitors in Non-small Cell Lung Cancer. Figure D.6: Defining biological sex based on sex chromosome complement. Scatterplot of first two principal components of Y chromosome gene expression in GTEx (top left), TCGA (top right), LGRC (bottom left) and GSE68465 (bottom right). Figure D.1: Sex difference in LGRC control lung samples within nonsmokers and smokers. Normalized enrichment scores (NES) from GSEA using KEGG pathways are shown for all pathways that have significant (adjusted p-value < 0.05) sex difference among either nonsmokers or smokers in LGRC. Pathways with higher targeting in male are marked blue and pathways with higher targeting in female are marked red. Green boxes highlight pathways associated with cell proliferation and brown boxes highlight pathways associated with environmental carcinogen metabolism. Figure D.2: Sex difference in tumor samples from the validation data GSE68465 within nonsmokers and smokers. Normalized enrichment scores (NES) from GSEA using KEGG pathways are shown for all pathways that have significant (adjusted p-value < 0.05) sex difference among either nonsmokers or smokers (in TCGA). Pathways with higher targeting in male are marked blue and pathways with higher targeting in female are marked red. Green boxes highlight pathways associated with cell proliferation and purple boxes highlight pathways associated with immune response. Figure D.5: Sex difference in immune and stromal cell composition in GTEx samples: nonsmokers (left) and smokers (right). Cell compositions are computed using "xcell", which derives cell composition proportion of 36 immune and stromal, along with three composite scores: immune score, stroma score and microenvironment score. The bubbleplot shows only those cells that are significantly (p-value < 0.05) different in proportion in male and female samples. Figure D.7: Violinplot representing the distribution of *XIST* expression in samples from GTEx (left) and TCGA (right), split by smoking history. Figure D.3: Heatmap representing the t-statistics corresponding to the sex differences in the ribosomal pathway from the limma model for the samples from both GTEx and TCGA, split by their smoking history. From the heatmap we observed that the targeting patterns of the 85 ribosomal genes included in our analysis is distinct among GTEx nonsmokers, compared to the other three groups. A positive value of the t-statistics corresponds to higher targeting in females and a negative value corresponds to higher targeting in males. Genes with higher targeting in males are marked blue and genes with higher targeting in females are marked red. Figure D.4: Sex difference in tumor samples from the validation data GIS031 within nonsmokers and smokers. Normalized enrichment scores (NES) from GSEA using KEGG pathways are shown for all pathways that have significant (adjusted p-value < 0.05) sex difference among either nonsmokers or smokers (in TCGA). A positive NES corresponds to higher targeting in females and a negative NES corresponds to higher targeting in males. Pathways with higher targeting in male are marked blue and pathways with higher targeting in female are marked red. Green boxes highlight pathways associated with cell proliferation and purple boxes highlight pathways associated with immune response. Table E.1: Distribution of Clinical Variables by Sex in GTEx. Table E.2: Distribution of Clinical Variables by Sex in TCGA. Table E.3: Distribution of Clinical Variables by Sex in LGRC. Table E.4: Distribution of Clinical Variables by Sex in GSE68465. Table E.5: Distribution of Clinical Variables among individuals who received chemotherapy versus those who did not, in TCGA dataset. F. Understanding Gene Regulatory Sex Differences in LUAD in East Asian Population.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Supplementary Material 6.

Supplementary Material 7.

Saha *et al. Biology of Sex Differences*   (2024) 15:62

Page 17 of 19

Availability of data and materials
Raw data to construct gene regulatory networks and other analysis were downloaded from open-source databases dbGap, Recount3, GEO, STRINGdb, CIS-BP and GDSC. Processed data are available from the corresponding author on reasonable request. The indegree matrices of the sample-specific gene regulatory networks are available in the GRAND database (https://grand.networkmedicine.org/downloads/) and can be found under the search term "sexDiffLUAD". Raw networks are available from the corresponding author on reasonable request. R codes for all downstream analysis are available on a GitHub public repository: https://github.com/Enakshi-Saha/Sex-Differences-Lung-Adenocarcinoma. A notebook describing differential targeting analysis on the TCGA data is available on Netbooks [58]: https://netbooks.networkmedicine.org.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA. [2]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA. [3]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA. [4]Department of Data Science, Dana-Farber Cancer Institute, Boston, MA 02115, USA.

References
1. Jemal A, Miller KD, Ma J, Siegel RL, Fedewa SA, Islami F, Devesa SS, Thun MJ. Higher lung cancer incidence in young women than young men in the United States. N Engl J Med. 2018;378(21):1999–2009.
2. Hellyer J, Patel M. Sex disparities in lung cancer incidence: validation of a long-observed trend. Transl Lung Cancer Res. 2019;8(4):543–5.
3. Fidler-Benaoudia MM, Torre LA, Bray F, Ferlay J, Jemal A. Lung cancer incidence in young women vs. young men: a systematic analysis in 40 countries. Int J Cancer. 2020;147(3):811–9.
4. Siegfried JM. Women and lung cancer: does oestrogen play a role? Lancet Oncol. 2001;2(8):506–13.
5. Mederos N, Friedlaender A, Peters S, Addeo A. Gender-specific aspects of epidemiology, molecular genetics and outcome: lung cancer. ESMO Open. 2020;5: e000796.
6. Yamamoto H, Sekine I, Yamada K, Nokihara H, Yamamoto N, Kunitoh H, Ohe Y, Tamura T. Gender differences in treatment outcomes among patients with non-small cell lung cancer given a combination of carboplatin and paclitaxel. Oncology. 2008;75(3–4):169–74.
7. Conforti F, Pala L, Bagnardi V, Viale G, De Pas T, Pagan E, Pennacchioli E, et al. Sex-based heterogeneity in response to lung cancer immunotherapy: a systematic review and meta-analysis. JNCI J Nat Cancer Inst. 2019;111(8):772–81.
8. Dresler CM, Fratelli C, Babb J, Everley L, Evans AA, Clapper ML. Gender differences in genetic susceptibility for lung cancer. Lung Cancer. 2000;30(3):153–60.
9. Omoto Y, Kobayashi Y, Nishida K, et al. Expression, function, and clinical implications of the estrogen receptor beta in human lung cancers. Biochem Biophys Res Commun. 2001;285(2):340–7.
10. Mollerup S, Jorgensen K, Berge G, et al. Expression of estrogen receptors alpha and beta in human lung tissue and cell lines. Lung Cancer. 2002;37(2):153–9.
11. Stapelfeld C, Dammann C, Maser E. Sex-specificity in lung cancer risk. Int J Cancer. 2020;146(9):2376–82.
12. McLemore T, Adelberg S, Liu M, et al. Expression of CYP1A1 gene in patients with lung cancer: evidence for cigarette smoke-induced gene expression in normal lung tissue and for altered gene regulation in primary pulmonary carcinomas. JNCI J Nat Cancer Inst. 1990;82(16):1333–9.
13. Patel J, Bach P, Kris M. Lung cancer in US women: a contemporary epidemic. JAMA. 2004;291(14):1763–8.
14. Li X, Wei S, Deng L, Tao H, et al. Sex-biased molecular differences in lung adenocarcinoma are ethnic and smoking specific. BMC Pulm Med. 2023;23(1):1–14.
15. Yuan Y, Liu L, Chen H, Wang Y, et al. Comprehensive characterization of molecular differences in cancer between male and female patients. Cancer Cell. 2016;29(5):711–22.
16. Glass K, Huttenhower C, Quackenbush J, Yuan G-C. Passing messages between biological networks to refine predicted interactions. PLoS ONE. 2013;8(5): e64832.
17. Kuijjer ML, Tung MG, Yuan G, Quackenbush J, Glass K. Estimating sample-specific regulatory networks. Iscience. 2019;14:226–40.
18. Lopes-Ramos C, Quackenbush J, DeMeo D. Genome-wide sex and gender differences in cancer. Front Oncol. 2020;23(10): 597788.
19. Glass K, Quackenbush J, Silverman E, Celli B, Rennard S, Yuan G-C, et al. Sexuallydimorphic targeting of functionally-related genes in COPD. BMC Syst Biol. 2014;8:1–17.
20. Lopes-Ramos CM, Kuijjer ML, Ogino S, Fuchs CS, DeMeo DL, Glass K, Quackenbush J. Gene regulatory network analysis identifies sex-linked differences in colon cancer drug metabolism. Can Res. 2018;78(19):5538–47.
21. Lopes-Ramos CM, Chen C-Y, Kuijjer ML, et al. Sex differences in gene expression and regulatory networks across 29 human tissues. Cell Rep. 2020;31(12):107795.
22. E. Saha, V. Fanfani and P. Mandros et. al, "Bayesian Optimized sample-specific Networks Obtained By Omics data (BONOBO)," *bioRxiv preprint*, pp. 2023–11, 2023.
23. Bamford S, Dawson E, Forbes S, Clements J, Pettett R, Dogan A, Flanagan A, Teague J, Futreal PA, Stratton MR, et al. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer. 2004;91(2):355–8.
24. Cheng D, Li J, Li S. CNOT 1 cooperates with LMNA to aggravate osteosarcoma tumorigenesis through the Hedgehog signaling pathway. Mol Oncol. 2017;11(4):388–404.
25. Lee J, Jung JH, Hwang J, et al. CNOT2 is critically involved in atorvastatin induced apoptotic and autophagic cell death in non-small cell lung cancers. Cancers. 2019;11(10):1470.

26. Shirai Y-T, Mizutani A, Nishijima S. CNOT3 targets negative cell cycle regulators in non-small cell lung cancer development. Oncogene. 2019;38(14):2580–94.

27. Zhang Y, Ma H. LncRNA HOXD-AS2 regulates miR-3681-5p/DCP1A axis to promote the progression of non-small cell lung cancer. J Thorac Disease. 2023;15(3):1289.

28. Grassadonia A, Sperduti I, Vici P, et al. Effect of gender on the outcome of patients receiving immune checkpoint inhibitors for advanced cancer: a systematic review and meta-analysis of phase III randomized clinical trials. J Clin Med. 2018;7(12):542.

29. Wagner A. Sex differences in cancer chemotherapy effects, and why we need to reconsider BSA-based dosing of chemotherapy. ESMO Open. 2020;5(5):e000770.

30. Rakshith HT, Lohita S, Rebello AP, et al. Sex differences in drug effects and/or toxicity in oncology. Curr Res Pharmacol Drug Discov. 2023;4: 100152.

31. Shtivelman E, Hensing T, Simon GR, Dennis PA, Otterson GA, Bueno R, Salgia R. Molecular pathways and therapeutic targets in lung cancer. Oncotarget. 2014;5(6):1392.

32. Ben Guebila M, Lopes-Ramos CM, Weighill D, et al. GRAND: a database of gene regulatory network models across human conditions. Nucleic Acids Res. 2022;50(D1):D610–21.

33. Giroux-Leprieur E, Costantini A, Ding VW, He B. Hedgehog signaling in lung cancer: from oncogenesis to cancer treatment resistance. Int J Mol Sci. 2018;19(9):2835.

34. Stewart DJ. Wnt signaling pathway in non-small cell lung cancer. JNCI J Nat Cancer Inst. 2014;106(1):djt356.

35. Zou B, Zhou X, Lai S, Liu J. Notch signaling and non-small cell lung cancer. Oncol Lett. 2018;15(3):3415–21.

36. Hoque MO, Brait M, Rosenbaum E, Poeta ML, Pal P, Begum S, Dasgupta S, et al. Genetic and epigenetic analysis of erbB signaling pathway genes in lung cancer. J Thorac Oncol. 2010;5(12):1887–93.

37. Mao K, Quipildor GF, Tabrizian T, Novaj A, Guan F, Walters RO, Delahaye F, Hubbard GB, Ikeno Y, Ejima K, et al. Late-life targeting of the IGF-1 receptor improves healthspan and lifespan in female mice. Nat Commun. 2018;9(1):2394.

38. Wright CJ, McCormack PL. Trametinib: first global approval. Drugs. 2013;73:1245–54.

39. Srinivas NR. Clinical pharmacokinetics of panobinostat, a novel histone deacetylase (HDAC) inhibitor: review and perspectives. Xenobiotica. 2017;47(4):354–68.

40. Mensah AA, Kwee I, et al. Novel HDAC inhibitors exhibit pre-clinical efficacy in lymphoma models and point to the importance of CDKN1A expression levels in mediating their anti-tumor response. Oncotarget. 2015;7(6):5059.

41. Richon VM, Sandhoff TW, et al. Histone deacetylase inhibitor selectively induces p21WAF1 expression and gene-associated histone acetylation. Proc Natl Acad Sci. 2000;97(18):10014–9.

42. Chen C, Saha E, DeMeo DL, Quackenbush J, Lopes-Ramos CM. Unveiling sex differences in lung adenocarcinoma through multi-omics integrative protein signaling networks. Cancer Res. 2024;84(6_Supplement):3490.

43. Wilks C, Zheng S, Chen F, et al. recount3: summaries and queries for large-scale RNA-seq expression and splicing. Genome Biol. 2021;22:323.

44. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, et al. The genotype-tissue expression (GTEx) project. Nat Genet. 2013;45(6):580–5.

45. The Cancer Genome Atlas Research Network, Weinstein J, Collisson E, et al. The cancer genome atlas pan-cancer analysis project. Nat Genet. 2013;45:1113–20.

46. Collado-Torres L, Nellore A, Kammers K, et al. Reproducible RNA-seq analysis using recount2. Nat Biotechnol. 2017;35:319–21.

47. Kim S, Herazo-Maya J, Kang D, et al. Integrative phenotyping framework (iPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes. BMC Genomics. 2015;16:924.

48. Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma. Gene expression–based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nat Med. 2008;14:822–7.

49. BenGuebila M, Wang T, Lopes-Ramos CM, Fanfani V, Weighill D, Burkholz R, Schlauch D, et al. The Network Zoo: a multilingual package for the inference and analysis of gene regulatory networks. Genome Biol. 2023;24(1):1–23.

50. Weirauch M, Yang A, Albu M, Cote A, Montenegro-Montero A, Drewe P, et al. Determination and inference of eukaryotic transcription factor sequence specificity. Cell. 2014;158:1431–43.

51. Szklarczyk D, Gable AL, Nastou KC, Lyon D, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021;2021(9):D605–12.

52. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43: e47.

53. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. BioRxiv. 2016; 2016.

54. Kanehisa M. The KEGG database. In: Silico' simulation of biological processes: novartis. New York: Wiley; 2002. p. 91–103.

55. 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.

56. Aran D, Hu Z, Butte A. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18:220.

57. Yang W, Soares J, Greninger P, Edelman EJ, et al. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 2012;41(D1):D955–61.

58. Ben Guebila M, Weighill D, Lopes-Ramos CM, et al. An online notebook resource for reproducible inference, analysis and publication of gene regulatory networks. Nat Methods. 2022;19(5):511–3.

59. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27(7):1017–8.

60. Herbst R, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet. 2016;387:1540–50.

61. Reck M, Rodríguez-Abreu D, Robinson A, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med. 2016;375:1823–33.

62. Rittmeyer A, Barlesi F, Waterkamp D, Park K, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet. 2017;389:255–65.

63. Carbone D, Reck M, Paz-Ares L, Creelan B, Horn L, et al. First-line nivolumab in stage iv or recurrent non-small-cell lung cancer. N Engl J Med. 2017;376:2415–26.

64. Antonia S, Villegas A, Daniel D, et al. Durvalumab after chemoradiotherapy in stage III non-small-cell lung cancer. N Engl J Med. 2017;377:1919–29.

65. Li X, Wei S, Deng L, Tao H, et al. Sex-biased molecular differences in lung adenocarcinoma are ethnic and smoking specific. BMC Pulm Med. 2023;23(1):99.

66. Sun X, Zhang T, Li M, Yin L, Xue J. Immunosuppressive B cells expressing PD-1/PD-L1 in solid tumors: a mini review. QJM. 2019;115:507–12.

67. Mueller F, Büchel B, Köberle D, et al. Gender-specific elimination of continuous-infusional 5-fluorouracil in patients with gastrointestinal malignancies: results from a prospective population pharmacokinetic study. Cancer Chemother Pharmacol. 2013;71:361–70.

68. Zhang Z, Xu L, Huang L, Li T, Wang JY, Ma C, Bian X, Ren X, Li H, Wang X. Glutathione S-transferase alpha 4 promotes proliferation and chemoresistance in colorectal cancer cells. Front Oncol. 2022;12: 887127.

69. Li W, Yue W, Zhang L, et al. Polymorphisms in GSTM1, CYP1A1, CYP2E1, and CYP2D6 are associated with susceptibility and chemotherapy response in non-small-cell lung cancer patients. Lung. 2012;190:91–8.

70. Rodriguez-Antona C, Ingelman-Sundberg M. Cytochrome P450 pharmacogenetics and cancer. Oncogene. 2006;25(11):1679–91.

71. Croker AK, Allan AL. Inhibition of aldehyde dehydrogenase (ALDH) activity reduces chemotherapy and radiation resistance of stem-like ALDH^hi CD44+ human breast cancer cells. Breast Cancer Res Treat. 2012;133:75–87.

72. Duan S, Gong B, Wang P, Huang H, Luo L, Liu F. Novel prognostic biomarkers of gastric cancer based on gene expression microarray: COL12A1, GSTA3, FGA and FGG. Mol Med Rep. 2018;18(4):3727–36.

73. Peng L, Zhuang L, Lin K, Yao Y, Zhang Y, Arumugam T, Fujii T, et al. Downregulation of GSTM2 enhances gemcitabine chemosensitivity of pancreatic cancer in vitro and in vivo. Pancreatology. 2021;21(1):115–23.

Saha *et al. Biology of Sex Differences*   (2024) 15:62

Page 19 of 19

74. Tibaldi C, Stasi I, Baldini E. Oncogene-addicted non-small-cell lung cancer in women: a narrative review of the importance of gender-related differences in treatment outcome. Precis Cancer Med. 2022;5:6.
75. Wagner AD, Oertelt-Prigione S, Adjei A, et al. Gender medicine and oncology: report and consensus of an ESMO workshop. Ann Oncol. 2019;30(12):1914–24.
76. Scheffler M, Holzem A, Kron A, et al. Co-occurrence of targetable mutations in Non-small cell lung cancer (NSCLC) patients harboring MAP2K1 mutations. Lung Cancer. 2020;144:40–8.

Publisher's Note

Enakshi Saha is a Research Associate at the Harvard T.H. Chan School of Public Health, Department of Biostatistics.

Marouen Ben Guebila is a Research Associate at the Harvard T.H. Chan School of Public Health, Department of Biostatistics.

Viola Fanfani is a Postdoctoral Fellow at the Harvard T.H. Chan School of Public Health, Department of Biostatistics.

Jonas Fischer was a Postdoctoral Fellow at the Harvard T.H. Chan School of Public Health, Department of Biostatistics.

Katherine H. Shutta is a Postdoctoral Fellow at the Harvard T.H. Chan School of Public Health, Department of Biostatistics.

Panagiotis Mandros was a Postdoctoral Fellow at the Harvard T.H. Chan School of Public Health, Department of Biostatistics.

Dawn L. DeMeo is Associate Professor of Medicine at Harvard Medical School and Associate Physician at Brigham and Women's Hospital.

John Quackenbush is a Professor of Computational Biology and Bioinformatics and Chair of the Department of Biostatistics at the Harvard T.H. Chan School of Public Health, Professor in the Channing Division of Network Medicine, and Professor at the Dana-Farber Cancer Institute.

Camila M. Lopes-Ramos is an Instructor in Medicine at the Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, and Instructor in Biostatistics at Harvard T.H. Chan School of Public Health.