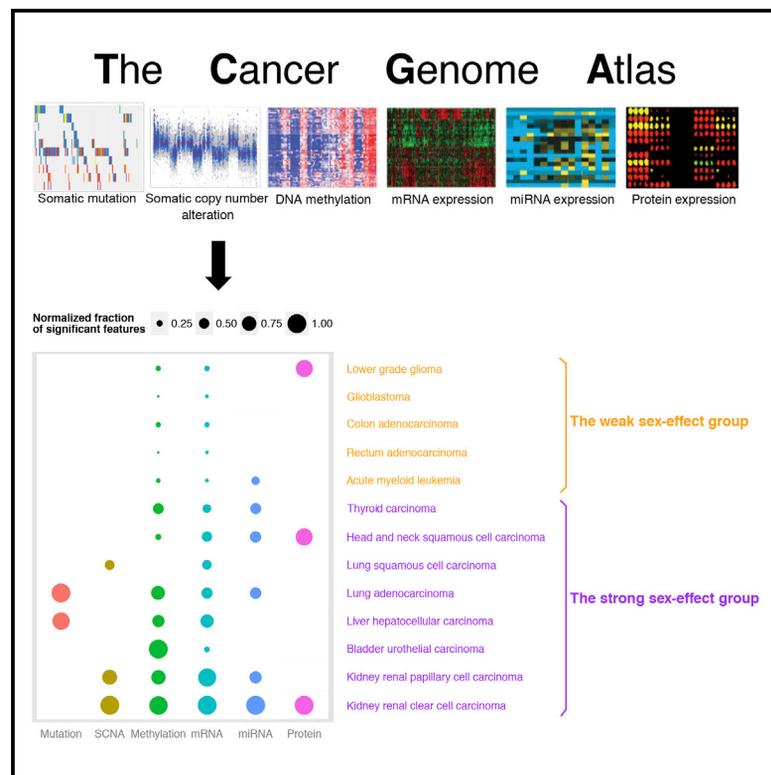


# Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients

## Graphical Abstract



## Authors

Yuan Yuan, Lingxiang Liu, Hu Chen, ..., Yongqian Shu, Liang Li, Han Liang

## Correspondence

hliang1@mdanderson.org

## In Brief

Yuan et al. perform a multidimensional analysis of molecular differences between male and female patients and classify cancer types into two groups based on sex-biased patterns. Many clinically actionable genes show sex-biased signatures in some tumor types, suggesting a need for sex-specific therapeutic strategies.

## Highlights

- A rigorous, pan-cancer analysis of sex effects on molecular profiles of patients
- Two sex-effect cancer groups showing distinct incidence and mortality profiles
- Extensive sex-biased gene expression signatures in some cancer types
- A considerable number of clinically actionable genes with sex-biased signatures



# Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients

Yuan Yuan,<sup>1,8</sup> Lingxiang Liu,<sup>1,2,8</sup> Hu Chen,<sup>1,3</sup> Yumeng Wang,<sup>1,3</sup> Yanxun Xu,<sup>4</sup> Huzhang Mao,<sup>5,6</sup> Jun Li,<sup>1</sup> Gordon B. Mills,<sup>7</sup> Yongqian Shu,<sup>2</sup> Liang Li,<sup>5</sup> and Han Liang<sup>1,3,7,\*</sup>

<sup>1</sup>Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>2</sup>Department of Oncology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210029, China

<sup>3</sup>Graduate Program in Structural and Computational Biology and Molecular Biophysics, Baylor College of Medicine, Houston, TX 77030, USA

<sup>4</sup>Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218, USA

<sup>5</sup>Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>6</sup>Department of Biostatistics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

<sup>7</sup>Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

<sup>8</sup>Co-first author

\*Correspondence: [hliang1@mdanderson.org](mailto:hliang1@mdanderson.org)

<http://dx.doi.org/10.1016/j.ccell.2016.04.001>

## SUMMARY

An individual's sex has been long recognized as a key factor affecting cancer incidence, prognosis, and treatment responses. However, the molecular basis for sex disparities in cancer remains poorly understood. We performed a comprehensive analysis of molecular differences between male and female patients in 13 cancer types of The Cancer Genome Atlas and revealed two sex-effect groups associated with distinct incidence and mortality profiles. One group contains a small number of sex-affected genes, whereas the other shows much more extensive sex-biased molecular signatures. Importantly, 53% of clinically actionable genes (60/114) show sex-biased signatures. Our study provides a systematic molecular-level understanding of sex effects in diverse cancers and suggests a pressing need to develop sex-specific therapeutic strategies in certain cancer types.

## INTRODUCTION

An individual's sex is a key factor affecting the risk of cancer development and management during his or her lifetime. This is not only because some cancer types are sex-specific (e.g., ovarian cancer in women and prostate cancer in men), but there are significant sex disparities in the incidence of cancer, tumor aggressiveness, prognosis, and treatment responses for many other cancer types (Branford et al., 2013; Cook et al., 2011; Dorak and Karpuzoglu, 2012; Molife et al., 2001; Pal and Hurria, 2010). However, the molecular basis for these observed disparities remains poorly understood. Previous studies have reported

some sex-related molecular patterns. For example, an elevated mutation rate of *EGFR* in female patients with non-small-cell lung cancer may contribute to enhanced response rates among female patients (Shepherd et al., 2005; Tam et al., 2006); and H3K27me3 demethylase UTX has been identified as a sex-specific tumor suppressor in T cell acute lymphoblastic leukemia (Van der Meulen et al., 2015). However, these studies on the sex effect have been limited to individual genes, single molecular data types, and single cancer lineages. Furthermore, previous cohort analyses of sex-affected molecular traits have been based on simple statistical tests without explicitly accounting for potentially confounding factors such as patient age,

### Significance

For many cancer types, men and women are very different in terms of susceptibility, survival, and mortality. But our knowledge about the differences between male and female cancer patients at the molecular level is very limited. This is a fundamental issue for cancer prevention and therapy but has not been investigated systematically. Through a rigorous, multidimensional analysis of sex-affected genes, we revealed a two-group molecular classification of cancer types (weak sex-effect group versus strong sex-effect group) and demonstrated that >50% of clinically actionable genes showed sex-biased molecular signatures in certain cancer types. Our study helps elucidate the molecular basis for sex disparities in cancer and lays a critical foundation for the future development of precision cancer medicine.

**Table 1. Summary of TCGA Cancer Types, Patient Samples, and Data Types Surveyed in This Study**

Cancer Type	Somatic Mutation		SCNA		DNA Methylation		mRNA Expression		miRNA Expression		Protein Expression	
	M	F	M	F	M	F	M	F	M	F	M	F
BLCA	130	44	135	46	142	47	141	47	142	46	83	32
COAD	162	144	218	197	142	125	222	205	207	189	174	153
GBM	171	98	322	208	65	48	94	53	NA	NA	123	82
HNSC	228	104	231	104	235	105	229	99	233	105	132	55
KIRC	275	148	314	168	192	102	321	177	311	165	294	144
KIRP	101	41	123	50	134	53	143	53	143	54	NA	NA
LAML	101	90	102	85	103	87	91	78	99	85	NA	NA
LGG	221	176	206	170	222	178	221	176	219	178	145	108
LIHC	75	53	76	51	91	60	89	57	90	57	NA	NA
LUAD	147	185	168	201	146	174	163	201	170	202	86	105
LUSC	111	42	274	97	184	65	273	97	260	92	128	44
READ	55	50	79	70	45	40	80	69	67	62	62	57
THCA	98	289	124	350	126	354	126	350	125	352	104	251

For each data type and each cancer type, the numbers of male and female patients available in the analysis are shown. NA, the data type in that cancer type was not included in the analysis. TCGA high-throughput characterization platforms are as follows. Somatic mutations: exome-sequencing data based on IlluminaGA/HiSeq automated DNA sequencing platform and SOLiD sequencing platform. Somatic copy-number alterations (SCNAs): Affymetrix Genome-Wide Human SNP Array 6.0. DNA methylation: Illumina Infinium Human DNA Methylation 450K Array. RNA expression: Illumina HiSeq 2000 RNA Sequencing V2. miRNA expression: Illumina Genome Analyzer/HiSeq 2000 miRNA sequencing platform. Protein expression: MD Anderson reverse-phase protein arrays. BLCA, bladder urothelial carcinoma; COAD, colon adenocarcinoma; GBM, glioblastoma multiforme; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; LAML, acute myeloid leukemia; LGG, brain lower grade glioma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; READ, rectal carcinoma; and THCA, thyroid carcinoma.

histological subtype, and tumor stage, which may introduce significant bias and misinterpretation. So far, a comprehensive characterization of molecular differences between male and female patients and related mechanisms across a broad range of human cancer types using a rigorous statistical approach has not been performed.

The availability of high-throughput molecular data over large, well-characterized patient sample cohorts of multiple cancer types through The Cancer Genome Atlas (TCGA) project provides an unprecedented opportunity to address this question (The Cancer Genome Atlas Research Network et al., 2013). Using these TCGA data, we performed a comprehensive, rigorous, pan-cancer analysis in order to address if and what the molecular-level differences are between the male and female cancer patients that have otherwise similar clinical and tumor characteristics.

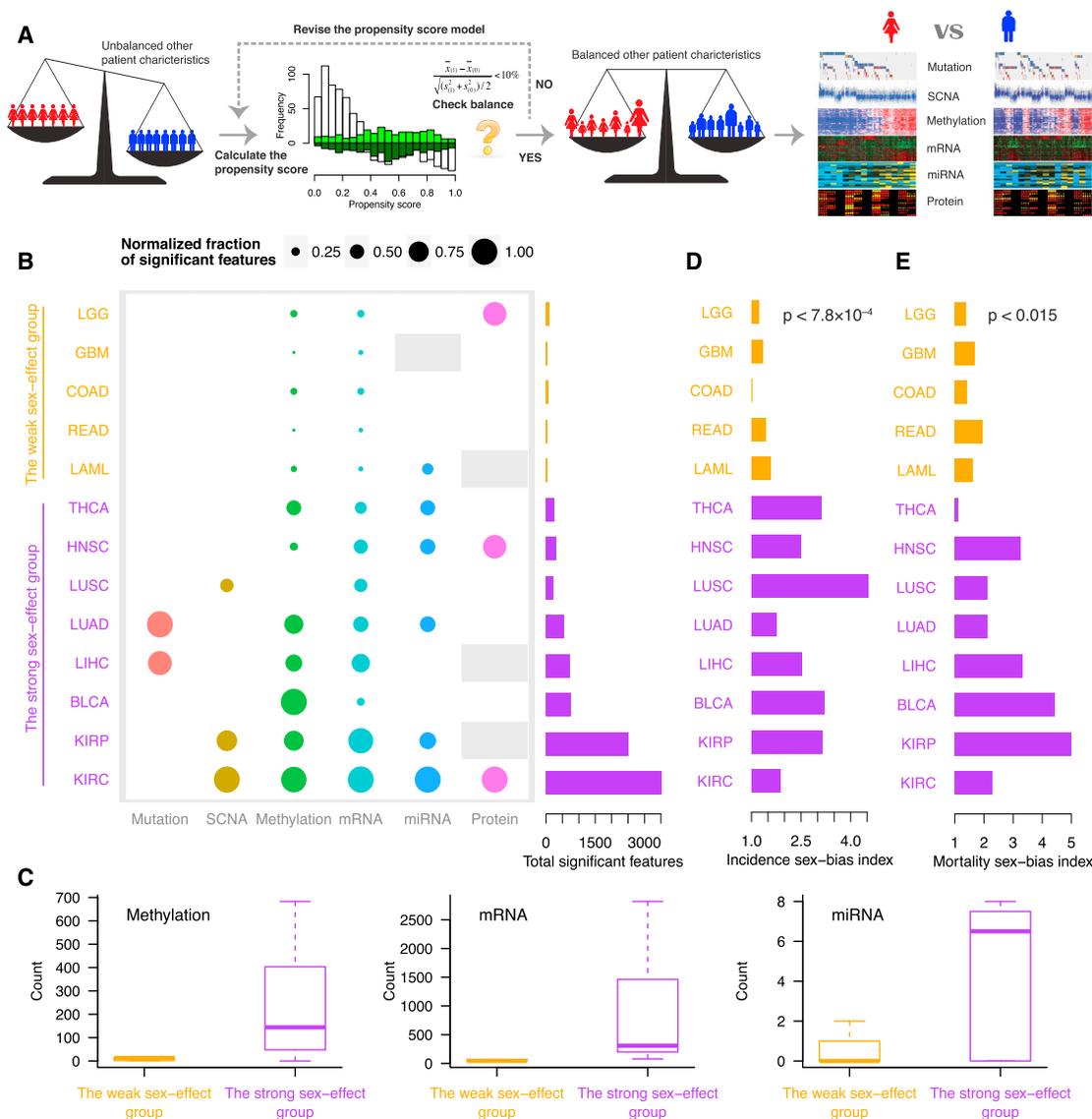
## RESULTS

### Overview of Multidimensional Sex-Affected Molecular Signatures across Cancer Types

We focused on 13 major TCGA cancer types with sufficient sample sizes ( $\geq 30$  for both male and female patients) for at least five out of six molecular data types (somatic mutations, somatic copy-number alterations [SCNAs], mRNA expression, DNA methylation, miRNA expression, and protein expression, Table 1). These cancer types include bladder urothelial carcinoma (BLCA), colon adenocarcinoma (COAD), glioblastoma multiforme (GBM), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma

(KIRP), acute myeloid leukemia (LAML), brain lower grade glioma (LGG), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), rectal carcinoma (READ), and thyroid carcinoma (THCA). It is important to note that these TCGA patient cohorts were not designed to study the sex effect, thus male and female patient groups in a cancer type are frequently different in other patient and tumor characteristics.

To identify molecular differences related to sex with appropriate controls for other factors that may bias findings (e.g., age, race, disease stage, and tumor purity, see potential confounders surveyed in Figure S1A), we employed an analytic approach based on the propensity score. Introduced in the early 1980s (Rosenbaum and Rubin, 1983), the propensity score is an important statistical tool for adjusting for confounding factors in observational studies, and has been widely used in clinical research, economics, and social sciences (D'Agostino, 1998; Ho et al., 2007; Imbens, 2004). Importantly, samples with the same propensity score have the same distribution of measured confounders, so balancing the confounders can be achieved by simply balancing the propensity score. As outlined in Figure 1A, we calculated the propensity scores, "reweighted" the samples in a cohort, and then compared the molecular features between the two balanced sex groups (Experimental Procedures). Our statistical simulations further confirmed that the propensity score method outperformed alternative methods in terms of both sensitivity and specificity (Figure S1B). With this approach, we identified significantly differential molecular features (genes) between female and male patients in the 13 cancer types (false discovery rate [FDR]  $\leq 0.05$ ). For the significant



**Figure 1. Overview of the Propensity Score Algorithm and the Sex-Affected Molecular Patterns across Cancer Types**

(A) An overview of the propensity score algorithm.

(B) Relative abundance of multidimensional sex-biased molecular signatures identified by the propensity score algorithm across cancer types ( $FDR \leq 0.05$ ). The fraction of significant features over total features was first calculated in each cancer type and then normalized across all cancer types. A gray box indicates that the specific data are not available for that cancer type. The bar plot shows the total number of significant features (by aggregating across all platforms) for each cancer type. The weak sex-effect and strong sex-effect groups are marked in orange and purple, respectively.

(C) The distribution of significant feature counts in the weak sex-effect group versus the strong sex-effect group (from left to right: DNA methylation, mRNA expression, and miRNA expression). The boundaries of the box mark the first and third quartile, with the median in the center, and whiskers extending to 1.5 interquartile range from the boundaries.

(D) The incidence sex-bias index for each cancer type.

(E) The mortality sex-bias index for each cancer type. The p values were calculated from Wilcoxon rank-sum tests to compare the two groups. See [Figure S1](#) and [Tables S1–S3](#).

feature set identified for a given data type and cancer type, we further confirmed its statistical significance by using permutation tests of randomly shuffling the sex labels of the patients ([Experimental Procedures, Figure S1C](#)). Focusing on the significant feature sets confirmed by the permutation tests, we examined the global patterns of sex-biased genes across different molecular types and found a clear separation among the cancer types

under survey. One group includes LGG, GBM, COAD, READ, and LAML, each of which shows a relatively small number of genes (44–104, mean 67) with a sex-biased pattern, which we therefore labeled the weak sex-effect group. The other group includes THCA, HNSC, LUSC, LUAD, LIHC, BLCA, KIRP, and KIRC, each of which shows much more extensive sex-biased molecular signatures (240–3,521, mean 1,112); we therefore

labeled it the strong sex-effect group (Figures 1B and Table S1). Indeed, no sex-biased somatically mutated genes or SCNAs were identified in any cancer of the weak sex-effect group; and the numbers of sex-biased genes at the mRNA, DNA methylation, and miRNA expression levels in this group were much lower than those in the strong sex-effect group (Figure 1C, Wilcoxon rank test, DNA methylation  $p < 0.015$ ; mRNA  $p < 0.0022$ ; miRNA  $p < 0.074$ ). Importantly, the sample sizes included in the analysis between these two cancer groups are similar, and so the observed distinct patterns cannot be attributed to the power to detect differences (Figure S1D).

Strikingly, compared with those in the weak sex-effect group, the cancer types in the strong sex-effect group show a higher incidence sex-bias index (defined on the basis of the ratio of new cases of female and male patients, Figures 1D and Table S2) and a higher mortality sex-bias index (defined on the basis of the ratio of the number of deaths among female and male patients, Figures 1E and Table S3). Furthermore, according to the National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology, a patient's sex has been suggested as a prognostic factor in five of the eight cancer types in the strong sex-effect group (i.e., LUSC, LUAD, HNSC, KIRC, and KIRP), but not in any cancer in the weak sex-effect group. We observed similar patterns based on other statistical cutoffs (e.g., FDR = 0.1 and 0.2, Figures S1E and S1F). Taken together, these results provide an overview of molecular differences between male and female cancer patients, and the distinct patterns of the two sex-effect groups are well aligned with the empirical observations of disease behaviors across cancer types.

### Sex-Biased Somatic Mutations and Copy-Number Alterations

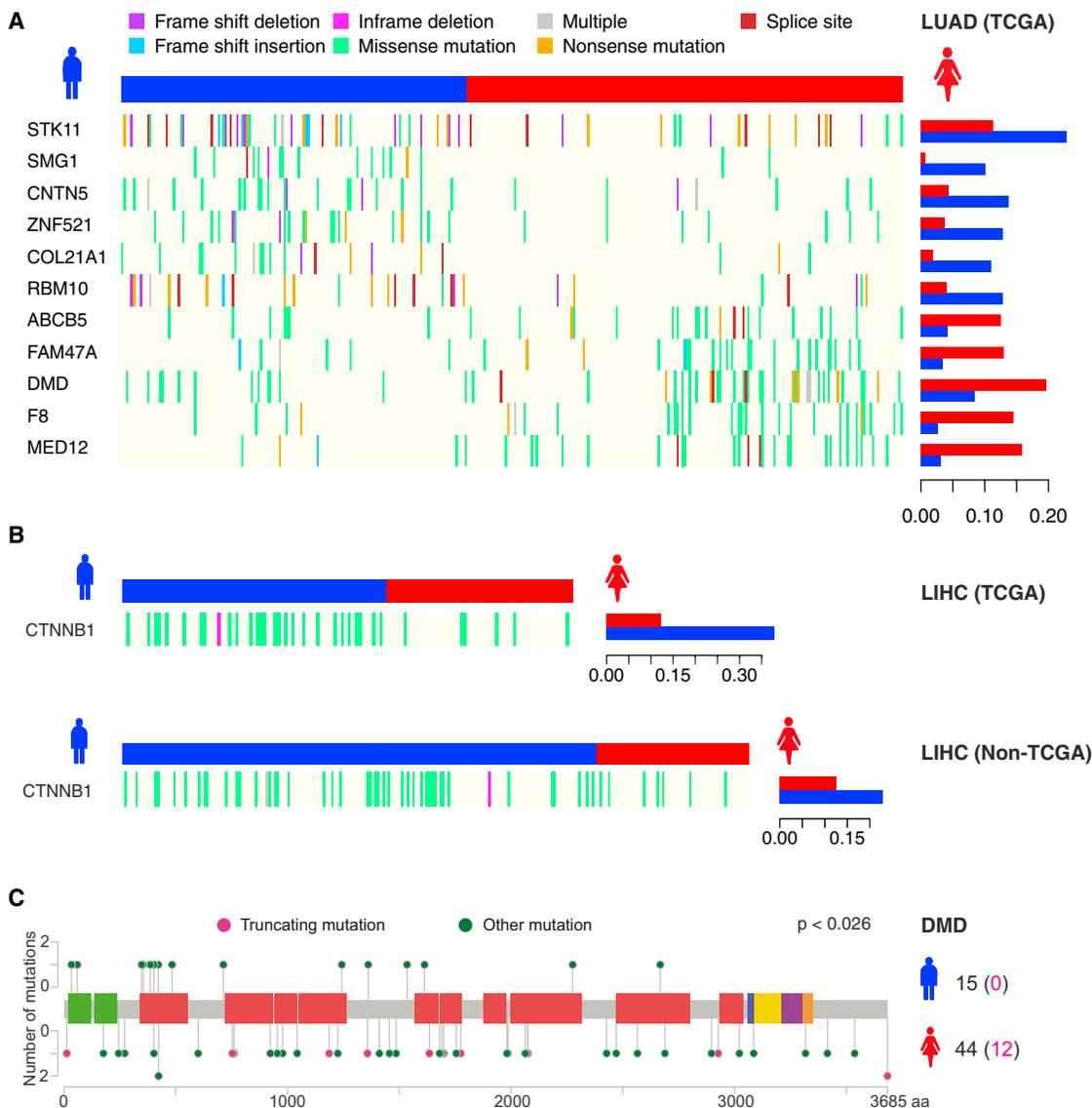
To identify sex-biased mutation patterns, we focused on highly mutated genes in each cancer type ( $\geq 5\%$  mutation frequency). The number of mutated genes under survey ranged from two in THCA to 650 in LUSC (Experimental Procedures). At FDR = 0.05, we identified 11 sex-biased genes in LUAD and one in LHC (Figures 2A and 2B). The most striking gene identified in LUAD was *STK11* (also known as *LKB1*), which encodes a major upstream kinase that activates the energy-sensing AMPK pathway and is frequently mutated in a variety of cancers (Jenne et al., 1998). Clinically, inactivating mutations in this gene may predict sensitivity to mTOR inhibitors, SRC inhibitors, and the metabolism drug phenformin in lung cancer (Carretero et al., 2010; Mahoney et al., 2009; Shackelford et al., 2013). Consistent with a previous analysis (The Cancer Genome Atlas Research Network, 2014), we found this gene to be more frequently mutated in males than in females, even after correcting for potential confounders (male versus female: 22.8% versus 11.3%,  $p < 6.9 \times 10^{-4}$ , FDR = 0.033). Another gene of interest in LUAD is *DMD*, which encodes a protein called dystrophin and is presumably responsible for Duchenne and Becker muscular dystrophies (Tennyson et al., 1995). The mutations in this gene were highly biased toward female patients (8.4% versus 19.6%,  $p < 3.7 \times 10^{-4}$ , FDR = 0.029). In particular, compared with the mutations in males, those in females had a greater tendency to be loss-of-function truncating mutations (Fisher's exact test,  $p < 0.026$ , Figure 2C). *EGFR*, a major therapeutic target in LUAD, showed a higher mutation frequency in female patients but did not reach

the FDR cutoff (9.8% versus 15.8%,  $p < 0.042$ , FDR = 0.28), which is consistent with previous reports (Marchetti et al., 2005; Schuette et al., 2015). The only sex-biased mutation gene we identified in LICH was *CTNNB1*, the activation of which could affect the sensitivity to EGFR inhibitors, PI3K inhibitors, AKT inhibitors, and WNT inhibitors (Anastas and Moon, 2013; Nakayama et al., 2014; Tenbaum et al., 2012). This gene is more frequently mutated in males (37.9% versus 12.2%,  $p < 1.2 \times 10^{-4}$ , FDR =  $7.4 \times 10^{-3}$ ), and we further confirmed this pattern in an independent sample cohort (Ahn et al., 2014) (22.9% versus 12.5%,  $p < 0.044$ , Figure 2C).

To identify sex-biased SCNAs, we focused on the most significant SCNAs identified by GISTIC (Mermel et al., 2011) in each cancer type. The number of region-based SCNAs (including both focal and arm-level amplifications/deletions) we surveyed ranged from 68 in KIRP to 122 in LUAD. At FDR = 0.05, we identified sex-biased SCNAs in LUSC, KIRP, and KIRC, all of which were in the strong sex-effect group. Figure 3 provides an overview of the statistical significance of sex-biased focal amplifications and deletions in these three cancer types, showing a total of 21 significant peaks (FDR  $\leq 0.1$ ). Notably, these sex-biased SCNAs cover quite a number of clinically actionable genes (as highlighted in Figure 3). Among them, two gene groups are of particular clinical interest. One group is related to the phosphoinositide 3-kinase (PI3K) pathway, which represents the signaling pathway most commonly activated in human cancer and has been under intensive clinical investigation (Liu et al., 2009), and related genes include *PIK3CA*, *MTOR*, *PTEN*, *NF1*, and *FBXW7*. In LUSC, an SCNA (17q11.2) harboring *NF1* is more frequently deleted in females, and the inactivation of this gene has been associated with sensitivity to mTOR inhibitors and resistance to MEK inhibitors (Janku et al., 2014; Nissan et al., 2014). In KIRP, the 4q34.3 deletion containing *FBXW7* occurs more frequently in females, and the deletion of this gene may affect the sensitivity to rapamycin treatment and antitubulin chemotherapeutics (Mao et al., 2008; Wertz et al., 2011). In KIRC, the amplicon 3q26 containing *PIK3CA* occurs more frequently in females, and *PIK3CA* activation has been reported to predict the sensitivity to PI3K/AKT/mTOR inhibitors (Janku et al., 2012); and the deletions of 1p36.23 (harboring *MTOR*) and 10q23.31 (harboring *PTEN*) are more prevalent in male patients. Another group is several therapeutic targets for cancer immunotherapy, which were detected in KIRC. *TNFRSF8* (CD30) and *CD52* are more frequently lost in males, and these two genes are the targets of Food and Drug Administration (FDA)-approved drugs for lymphoma and B cell chronic lymphocytic leukemia, respectively (Buggins et al., 2002; Younes et al., 2013). The deletion involving *PDCD1* (PD-1) shows a similar bias; this gene represents an immune checkpoint and has been a major focus in the development of immunotherapy (Pardoll, 2012).

### Sex-Biased Gene Expression Signatures

To characterize the sex-biased gene expression signatures in a comprehensive manner, we performed analyses on RNA expression (~20,000 genes, including ~17,000 protein-coding genes and ~3,000 noncoding genes), DNA methylation (~16,000 protein-coding genes), miRNA (~500), and protein expression (191 proteins and phosphorylated proteins). For RNA expression, the number of sex-biased genes in the weak sex-effect group was



**Figure 2. Sex-Biased Somatic Mutation Signatures**

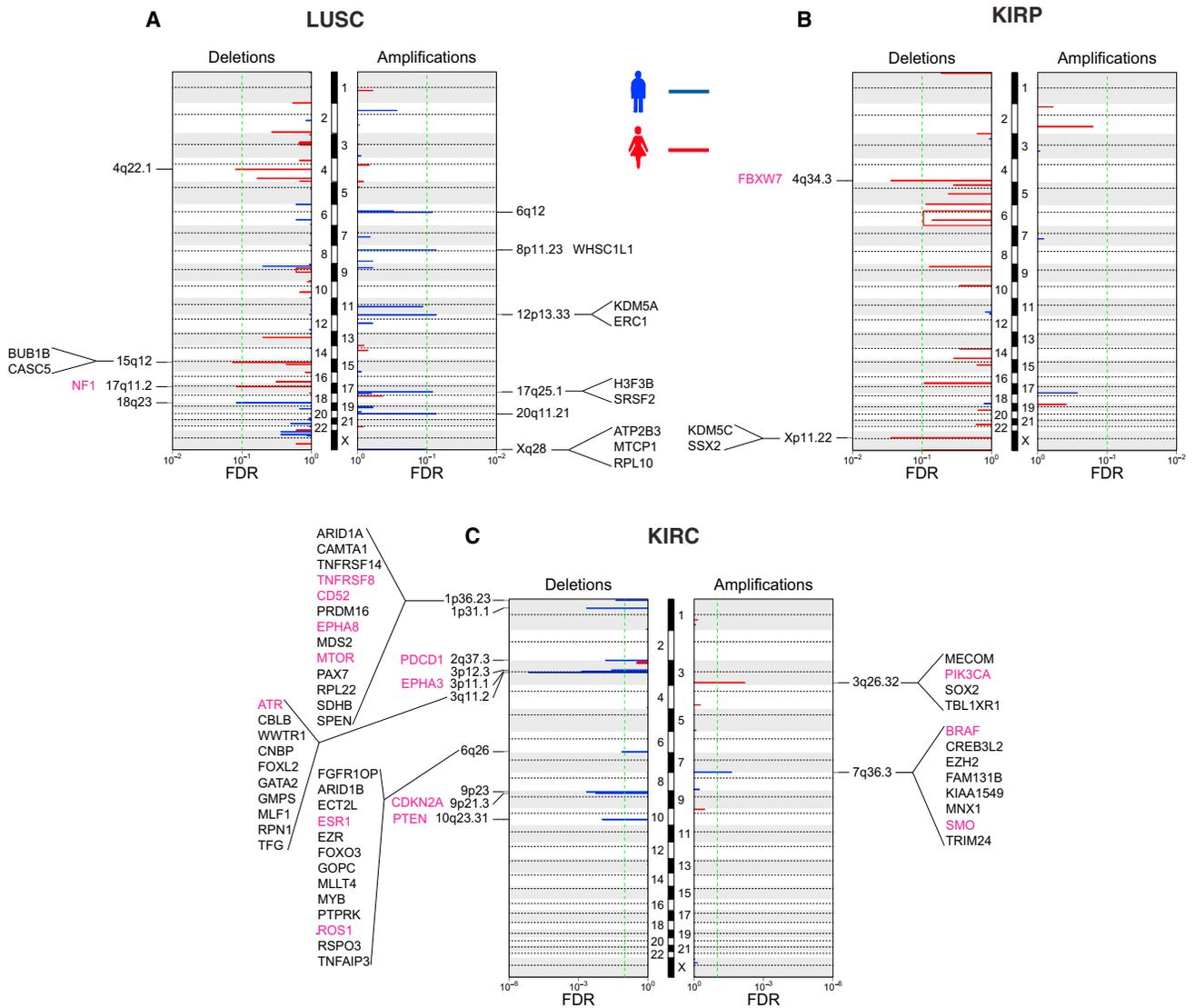
(A and B) Overview of the genes with a sex-biased mutation signature in (A) LUAD and (B) LIHC (FDR  $\leq 0.05$ ). Samples are displayed as columns with the sex label on the top, and different colors indicate different types of somatic mutations. The bar plots next to the heatmaps show the recalibrated mutation frequencies after propensity score weighting.

(C) The lollipop plot shows the sex-biased patterns for *DMD* in LUAD, with truncating mutations in magenta and other mutations in green. The numbers in parentheses summarize the number of truncating mutations, and the p value was calculated with Fisher's exact test.

very limited (Figure 1C); while the number of sex-biased genes in the strong sex-effect group was much higher (ranging from 79 in BLCA to 2,819 in KIRC, FDR  $\leq 0.05$ ), up to 14% of the whole gene set under survey. As expected, we found that the sex-biased genes were significantly enriched in the sex chromosomes (i.e., chrX and chrY); and, in particular, the vast majority (88%) of the sex-biased genes in at least four cancer types come from these two chromosomes (Figure S2A, Fisher's exact test,  $p = 2.2 \times 10^{-16}$ ). For comparison, we performed a similar analysis of the mRNA expression data from related normal tissue samples of the five cancer types in the strong sex-effect group (Table S4). Although much fewer sex-biased genes were detected in the normal samples (likely due to the much smaller

sample sizes), we observed the same enrichment of sex-biased genes in the sex chromosomes. One of the most commonly identified genes is *XIST*, a major effector of chromosome X inactivation; and its role in cancer has been extensively studied (Ganesan et al., 2002; Vincent-Salomon et al., 2007). In parallel, we found many more genes with a sex-biased DNA methylation pattern in the strong sex-effect group than in the weak sex-effect group (Figure 1C). We also identified sex-biased miRNA genes in six cancer types, five of which are in the strong sex-effect group.

Focusing on the eight cancer types in the strong sex-effect group, we further examined the genes identified by RNA expression (FDR  $\leq 0.05$ ), and found that, in all the cancer types, the sex bias observed at the mRNA level of a gene tended to be the



**Figure 3. Sex-Biased Somatic Copy-Number Alterations**

(A–C) The genome-wide, sex-biased focal amplification/deletion patterns in (A) LUSC, (B) KIRC, and (C) KIRC. The male-biased SCNA peaks are shown in blue, and the female-biased ones are in red. The significant SCNA regions ( $FDR \leq 0.1$ , indicated by the vertical green dotted lines) harboring important cancer genes are annotated, and the clinically actionable genes are highlighted in purple.

opposite of that at its DNA methylation level. This is consistent with the established role of DNA methylation in gene regulation: hypermethylation leads to gene silencing, while hypomethylation results in the up-regulation of gene expression (Figure 4A). To gain insight into the global patterns of the sex effect on gene expression, we performed a gene-set-enrichment analysis (GSEA) given the gene ranks according to the sex bias, and identified the affected pathways. In general, we observed biologically sensible, contrasting sex-bias patterns at the mRNA and DNA methylation levels (Figure 4B). We obtained similar results after excluding the genes in the sex chromosomes (Figure S2B). Thus, both gene-based and gene-rank-based pathway-enrichment analyses indicated that the sex-biased mRNA expression patterns in the strong sex-effect cancers are partially the result of the corresponding sex-biased DNA methylation.

Combining the analyses on mRNA and DNA methylation, we identified several themes among the sex-affected pathways. The first group relates to the immune response, including allograft rejection, IL2 and STAT5 signaling, IL6, JAK, and STAT3 signaling, inflammatory responses, interferon alpha response, interferon gamma response, and TNF- $\alpha$  signaling and complement. These enrichments are well aligned with the long-standing observation that females and males often mount significantly different immune responses (Purtulo and Sullivan, 1979; Weinstein et al., 1984). The second group relates to apoptosis and the cell cycle, including E2F targets, the G2/M checkpoint, mitotic spindle, and Myc targets. The targets of E2F transcription factors play a major role during the G1/S transition and, interestingly, some target genes encode differentiation factors that are transcribed in developmentally regulated and sex-specific

patterns (Dimova et al., 2003). The third group is several metabolism-related pathways such as angiogenesis, bile acid metabolism, fatty acid metabolism, glycolysis, and xenobiotic metabolism. Notably, sex-related metabolic differences and hormonal regulation have been reported in several studies (Drolz et al., 2014; Mittelstrass et al., 2011). Finally, DNA repair and P53 pathways also show sex-biased expression signatures in several cancer types including HNSC, KIRC, and LIHC.

As for protein expression, we found abundant sex-biased protein expression signals in HNSC and KIRC. Interestingly, 12 (of the 15) sex-biased proteins identified in HNSC or 18 (of the 25) sex-biased proteins identified in KIRC form well-connected regulatory networks (Figures 4C and 4D); SRC and MAPK proteins take a central position in both networks. SRC plays a key role in regulating a variety of cellular-signaling transduction pathways, and the frequent activation of the SRC kinase pathway has been observed in many cancer types, especially in metastatic diseases (Dehm and Bonham, 2004). Thus, targeted inhibition of SRC signaling has been suggested as an effective therapeutic strategy and has been under intensive clinical investigation in several cancers, including renal cell cancer (Araujo and Logothetis, 2010; Suwaki et al., 2011). Indeed, the factor of female sex has been reported to predict the response of imatinib, an inhibitor of BCR-ABL tyrosine kinase that also affects the SRC/MAPK pathway, in patients with chronic myeloid leukemia (Deininger et al., 2003; Valeyrie et al., 2003). This finding may be relevant to the female-biased expression pattern observed here.

### Sex-Biased Molecular Signatures of Clinically Actionable Genes

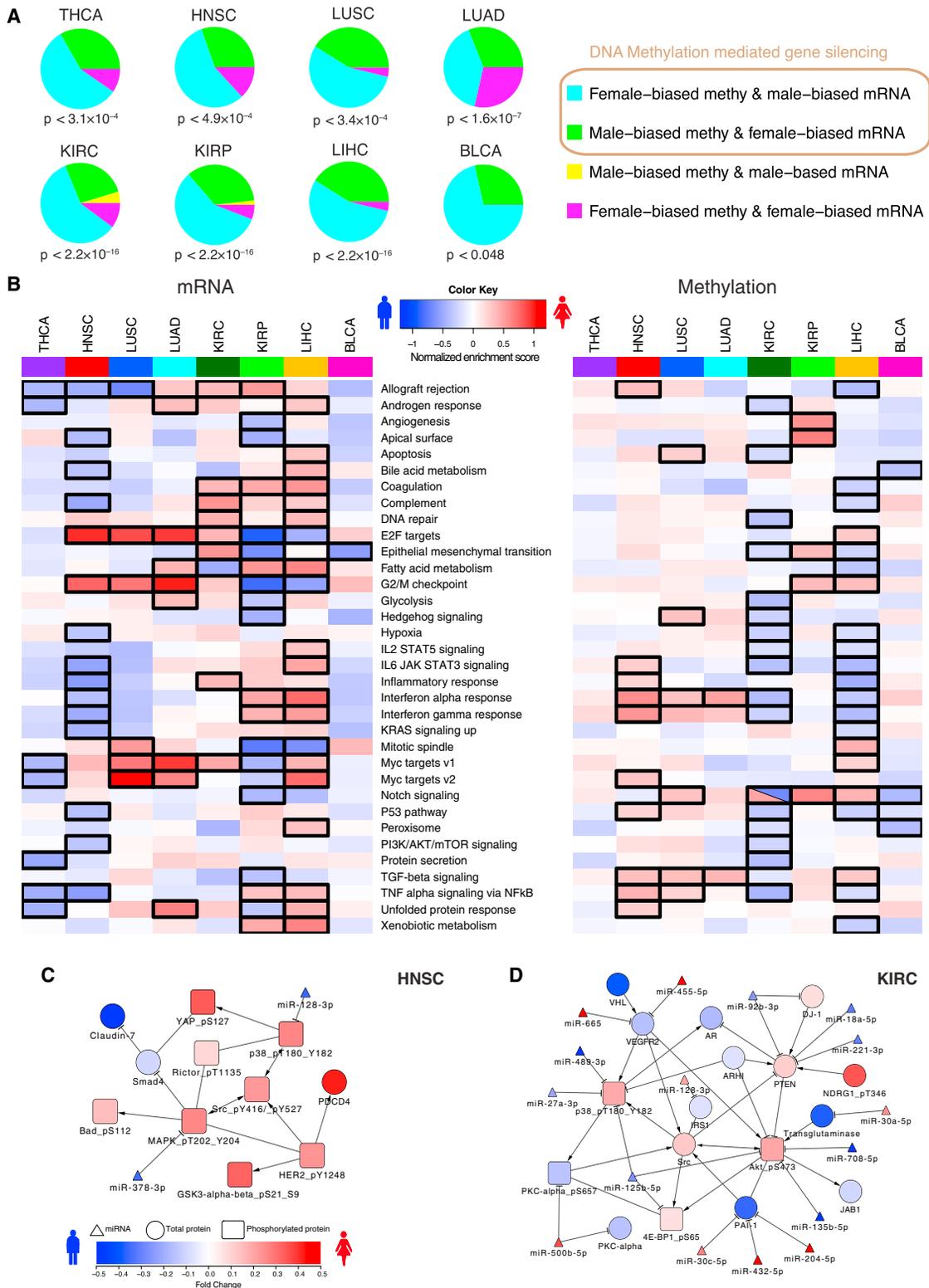
To investigate the clinical implications of the sex-biased molecular signatures, we focused on a set of clinically actionable genes, which includes 86 therapeutic targets of FDA-approved drugs or their associated predictive marker genes. Across the various molecular dimensions we examined (Figure S3), we found that 60 genes are associated with at least one type of sex-biased signature, almost all of which were identified in seven of the eight cancer types in the strong sex-effect group, ranging from two in LUAD/LUSC to 32 in KIRC (Figure 5). Among these genes, quite a few showed sex-biased signatures in the cancer type for which their relevant drugs are being used in clinical practice, which is of particular importance. For example, *EGFR*, arguably the most important therapeutic target in LUAD, shows female-biased mRNA expression, which may contribute to a higher response rate in female patients (Shepherd et al., 2005). *TOP2B* shows male-biased DNA methylation in BLCA, and the relevant drug valrubicin is being used as an intravesical therapy for BCG-refractory carcinoma in situ of the urinary bladder. Valrubicin can increase the risk of heart failure, but this side effect can be suppressed by tamoxifen, an agonist of estrogen (Zhang et al., 2012), which suggests that the innately distinct background level of sex hormone (e.g., estrogen levels) between male and female patients may lead to different drug responses or efficacy. Across cancer types, the two kidney cancers (KIRC and KIRP) contain the most clinically actionable genes, with a KIRP signature including *FBXW7*, *FGFR1*, *RET*, and *TSC2*, and a KIRC signature including *FGFR3*, *AKT*, *TSC2*, and *KIT*. These results highlight the clinical importance of sex-biased molecular signatures.

### DISCUSSION

Although the significance of the sex effect in cancer incidence, prognosis, and treatment responses has long been recognized in the literature, its molecular basis has largely remained elusive. Our study represents a comprehensive and well-controlled analysis that focuses on the molecular differences between male and female patients across a broad range of cancer types, and systematically catalogues the molecular signatures related to the sex effect from DNA to RNA to protein. Using the propensity score algorithm, we controlled the potential confounders (both patient characteristics such as age and smoking status, and tumor characteristics such as stage, histology subtype, and tumor purity) whenever possible in the analysis. Thus, for the TCGA datasets we assessed, the detected molecular differences between the two sex-effect groups could not be attributed to these potential confounders. Based on multidimensional molecular signatures, we defined two distinct cancer groups (weak sex-effect versus strong sex-effect): the cancer types in the weak group contain a very limited number of sex-biased genes and are associated with more balanced incidence and mortality ratios; whereas cancers in the strong group show extensive sex-biased molecular signatures and are associated with more distorted incidence and mortality ratios. This molecular classification of cancer types we put forward will help to achieve a molecular-level understanding of how the sex factor affects the behavior of different cancer types.

Given the widespread sex-biased gene expression signatures in the strong sex-effect group, it would be interesting to identify to what extent the observed sex bias is specific to cancer samples and elucidate the contributing factors. Our analysis based on the mRNA expression data of related TCGA normal samples detected much fewer sex-biased genes, suggesting the sex bias might be amplified during the tumorigenesis process. However, this observation should be interpreted with caution, because (1) the so-called normal tissues actually consist of quite distinct cell types from the corresponding tumor samples (e.g., the proportion of epithelial cells), which may confound the observed tumor-normal differences; and (2) the sample size of normal tissues is much smaller than that of the corresponding tumor samples, which may limit the detection power. Thus, further efforts are required to elucidate the relative contributions of various factors (e.g., sex chromosomes, hormones, and tumorigenesis) to the observed sex-biased gene expression signatures in cancer samples.

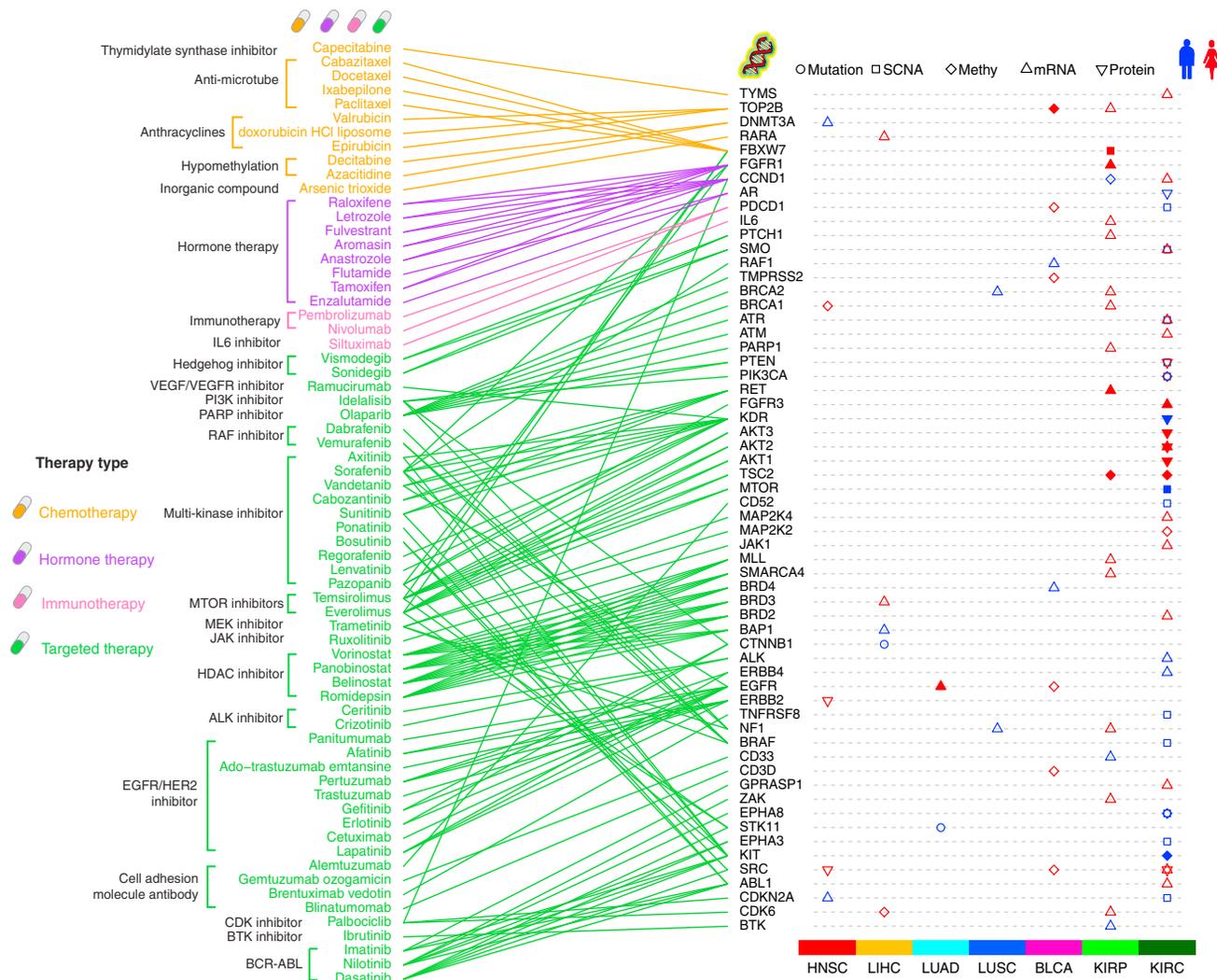
Our study reports 53% of clinically actionable genes (i.e., therapeutic targets and biomarkers) with a sex-related molecular pattern, and this proportion is significantly higher than that for all protein-coding genes (39%). However, this enrichment may not be surprising. Because most genes were not or rarely mutated or gained/lost in cancer samples, we could only examine the sex effect for highly mutated genes or frequent SCNAs. These genes or regions are enriched for cancer driver genes, and driver genes tend to be clinically actionable genes. As for protein expression data, those protein markers in the profiling platform were selected to represent clinically actionable genes. When we focused on mRNA and DNA methylation data, the two relatively unbiased platforms, and repeated the analysis, we observed the sex-biased signatures in 40% (46/114) of



**Figure 4. Sex-Biased Gene Expression Signatures**

(A) Pie charts showing the distributions of genes with both significant sex-biased mRNA expression and DNA methylation patterns for the eight cancer types in the strong sex-effect group. The p values were calculated from Fisher’s exact test to assess the association between the sex-biased methylation and mRNA expression patterns.

(legend continued on next page)



**Figure 5. Sex-Biased Molecular Signatures of Clinically Actionable Genes**

The mapping between FDA-approved drugs and their related clinically actionable genes (left) and the observed sex bias of these clinically actionable genes across cancer types (right). Different symbol shapes indicate different types of molecular signatures, and the filled shapes indicate that the gene is a therapeutic target of clinical practice in the corresponding cancer type. See Figure S3.

clinically actionable genes, which is no longer statistically significant. Such a systematic identification of sex-biased signatures for clinically actionable genes has crucial clinical implications. Currently, male and female patients with many cancer types are often treated in a similar way without explicitly considering the factor of sex. While this practice may be appropriate for the cancer types in the weak sex-effect group, special consideration should be given to those in the strong sex-effect group in terms of both drug development and clinical practice. For a therapeutic target with a strong sex-biased signature, sex-specific clinical trials may be more likely to succeed. For example, *SRC* appeared to have a much higher protein expression level in

females than in males with HNSC, but two recent dasatinib-driven clinical trials in this disease failed (Brooks et al., 2011; Fury et al., 2011), which might be due to the small proportion of female patients recruited in these studies (4/15 and 2/9, respectively). Our results thus provide a valuable starting point from which sex-specific effects should be explicitly considered in future clinical investigations. In clinical practice, even when the molecular data for a specific drug target are not available for a patient, it would be helpful to use the sex-biased molecular signatures identified as prior knowledge when making a choice among different treatment options. Since TCGA clinical information may not be complete and rigorously annotated, future

(B) The biological pathways identified by GSEA based on the sex-biased gene ranks of mRNA expression (left) and DNA methylation (right). Boxes highlight the statistically significant enriched pathways (mRNA: FDR ≤ 0.05; DNA methylation: FDR ≤ 0.2), and enrichment for female-biased and male-biased genes are shown in red and blue, respectively.

(C and D) The gene regulatory networks formed by the proteins with a sex-biased protein expression level (FDR ≤ 0.05) and their potential miRNA regulators in (C) HNSC and (D) KIRC. See Figure S2 and Table S4.

studies on this topic would require analyses on additional patient cohorts with more carefully annotated clinical variables, as well as the efforts of assessing the clinical utility of the sex-biased signatures identified.

## EXPERIMENTAL PROCEDURES

### Propensity Score Algorithm

We obtained TCGA patient and tumor characteristics (e.g., sex, age at diagnosis, smoking status, tumor stage, and histology subtype) for the 13 cancer types from the TCGA data portal (<https://tcga-data.nci.nih.gov/tcga/>), and the tumor purity data from Synapse: syn3242754. We obtained various types of TCGA molecular data as described in subsequent sections. Given the patient and tumor variables in the sample cohort for a specific molecular data type, we first calculated the propensity score based on “sex” using logistic regression. Then we employed the matching weight scheme (Li and Greene, 2013) to re-weight the samples based on the calculated scores. This step balanced the propensity scores and ultimately the covariates. The design followed a strict checking loop so that the propensity score model could be revised continuously until all covariates were balanced between the male and female groups (i.e., standardized differences <10%). After completing the above procedure, we compared the molecular data between the two balanced groups by supplying the weight vector calculated from the balancing step to a linear regression model using sex as the sole independent variable, and quantified the relative fold-change and corresponding statistical significance (e.g., raw  $p$  and FDR) of the sex effect. For each data type and each cancer type, we identified a significant feature (gene) set at  $FDR \leq 0.05$ . To further ensure that the signals detected were above the level of random noise, we performed permutation tests by randomly shuffling the sex label of the patients (while the other variables remained the same) and repeated the propensity score balancing/calculation procedure on the permuted data. The above procedure was conducted independently 100 times. We then compared the hit number in a significant feature set inferred from the original dataset with those from the permuted datasets to assess whether the signals we observed from the original data were true (caused by sex) or due to random noise. Only the significant feature sets showing statistical significance ( $p \leq 0.05$ ) in the permutation tests were retained for further analysis (Figure S1C).

### Patient Statistical Data Analysis

We obtained the incidence and mortality data from the literature (Lipworth et al., 2014; Ostrom et al., 2014; Siegel et al., 2015) and the Surveillance, Epidemiology, and End Results Program of the National Cancer Institute (<http://seer.cancer.gov/data/>). For both incidence and mortality rates, the sex-bias index was defined as the maximum (female-to-male ratio, male-to-female ratio). The Wilcoxon sum rank test was used to compare the sex-bias indexes between the weak sex-effect and strong sex-effect groups. Cancer prognostic factor information was obtained from the NCCN Clinical Practice Guidelines in Oncology ([www.nccn.org/professionals/physician\\_gls](http://www.nccn.org/professionals/physician_gls)).

### Analysis of Somatic Mutation Data

We obtained the mutation data (MAF files) from Firehose (<http://gdac.broadinstitute.org>) (2015 April) and retained only non-silent mutations for analysis. To prevent the potential bias introduced by ultramutated samples, we filtered out the samples with >1,000 mutations in their exomes. We focused on the non-silent mutations with  $\geq 5\%$  mutation frequency in a patient cohort because of their potential biological significance and detecting power in the analysis. We applied the propensity score algorithm to identify the mutated genes that show significant differences between male and female patients at  $FDR = 0.05$  and permutation test  $p = 0.05$ . We obtained an independent cohort of liver cancer (Ahn et al., 2014), and performed the same procedure as for the TCGA data. For selected genes of interest, we extracted mutation information from the MAF file and generated the lollipop plot using MutationMapper from cBioPortal (Cerami et al., 2012). Truncating mutations referred to those involving frame shift, nonsense and splicing sites. Fisher’s exact test was used to compare truncating versus other mutation patterns between male and female patients.

### Analysis of Somatic Copy-Number Alterations

We obtained significant SCNAs (both focal- and arm-level) from Firehose (2014 April). The propensity score algorithm was applied to identify the cancer types with a sex-biased SCNA at  $FDR = 0.05$  and permutation test  $p = 0.05$ . The sex bias was determined according to the relative levels of SCNAs in male and female patients, and their nature (i.e., whether it was an amplification peak or deletion peak). The cancer genes were annotated (Futreal et al., 2004). Clinically actionable genes are defined below.

### Analysis of mRNA Expression and DNA Methylation Data

We obtained normalized mRNA expression data based on RNA-seq (RNA-Seq by expectation maximization [RSEM]) from the TCGA data portal. The propensity score algorithm was applied to the  $\log_2$ -transformed RSEM to identify the genes that show significant differences between male and female patients at  $FDR = 0.05$  and permutation test  $p = 0.05$ . To identify sex-biased pathways, we performed GSEA (Subramanian et al., 2005) based on the full set of genes, ranked on the basis of the female-biased or male-biased mRNA expression fold-change and statistical significance, and detected significant pathways at  $FDR = 0.05$ . We performed a similar analysis on gene expression data of normal tissue samples from five cancer types.

We obtained DNA methylation 450K data from the TCGA data portal. Since multiple methylation probes can be mapped to a single gene, we first generated a one-to-one gene-methylation probe mapping by preserving the methylation probes that are most negatively correlated with the corresponding gene expression. We then applied the propensity score algorithm to the re-annotated methylation data, performed the same GSEA as for the mRNA expression data, and detected significant pathways at  $FDR = 0.2$ . In addition, to elucidate the regulatory mechanism of sex-biased gene expression patterns, for the genes with a sex-biased mRNA signature ( $FDR \leq 0.05$ ), we examined whether their methylation patterns were significantly sex-biased ( $p \leq 0.05$ ) and used Fisher’s exact test to assess the concordance of their directions (i.e., whether it was a male-biased or female-biased signature).

### Protein and miRNA Expression Analyses

We obtained the protein expression data from The Cancer Proteome Atlas (Li et al., 2013) and the miRNA expression data (in reads per million) from Firehose (2014 October). The propensity score algorithm was applied to identify the proteins/miRNAs that showed significant differences between male and female patients at  $FDR = 0.05$  and permutation test  $p = 0.05$ . Among these candidate miRNAs, we further identified potential miRNA regulators for these sex-biased proteins using two criteria: (1) the mature miRNA has the identified sex-biased protein genes as either experimentally validated targets from miRTarBase (Hsu et al., 2011), or computationally predicted targets from three well-established miRNA-target prediction databases TargetScan, miRanda and miRDB (John et al., 2004; Lewis et al., 2003; Wong and Wang, 2015); and (2) the candidate miRNA shows the opposite sex bias as that of the protein.

### Analysis of Clinically Actionable Genes and Drugs

We defined clinically actionable genes as FDA-approved therapeutic targets and their relevant predictor markers (Van Allen et al., 2014). We obtained the hematology/oncology (cancer) drugs and prescription information from the website (<http://www.fda.gov/Drugs/InformationOnDrugs/>) during the period of 1995–July 2015.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.ccell.2016.04.001>.

## AUTHOR CONTRIBUTIONS

H.L. conceived and supervised the project. Y.Y., L. Liu, and H.L. designed and performed the research. H.C., Y.W., Y.X., H.M., J.L., G.B.M., Y.S., and L. Li contributed to the data analysis. Y.Y., L. Liu, and H.L. wrote the manuscript, with input from all other authors.

## ACKNOWLEDGMENTS

We gratefully acknowledge contributions from TCGA Research Network and TCGA Pan-Cancer Analysis Working Group. This study was supported by the NIH (CA175486 and CCSG grant CA016672, H.L.); the R. Lee Clark Fellow Award from The Jeanne F. Shelby Scholarship Fund (H.L.); a grant from the Cancer Prevention and Research Institute of Texas (RP140462, H.L.); the Lorraine Dell Program in Bioinformatics for Personalization of Cancer Medicine (H.L.); by the National Natural Science Foundation of China (81472782, L.Liu), Natural Science Foundation of Jiangsu Province (BK20141491, L.Liu), Six Talent Peaks Foundation of Jiangsu Province (2012-WS-026, L.Liu), and “333” Talents Project of Jiangsu Province (L.Liu). We thank the MD Anderson high-performance computing core facility for computing, LeeAnn Chastain for editorial assistance, and Eli Van Allen for valuable comments. G.B.M. has sponsored research support from AstraZeneca and is on the Scientific Advisory Board for AstraZeneca, ImmunoMet, Nuevolution and Precision Medicine.

Received: November 7, 2015

Revised: February 15, 2016

Accepted: April 1, 2016

Published: May 9, 2016

## REFERENCES

- Ahn, S.M., Jang, S.J., Shim, J.H., Kim, D., Hong, S.M., Sung, C.O., Baek, D., Haq, F., Ansari, A.A., Lee, S.Y., et al. (2014). Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology* *60*, 1972–1982.
- Anastas, J.N., and Moon, R.T. (2013). WNT signalling pathways as therapeutic targets in cancer. *Nat. Rev. Cancer* *13*, 11–26.
- Araujo, J., and Logothetis, C. (2010). Dasatinib: a potent SRC inhibitor in clinical development for the treatment of solid tumors. *Cancer Treat. Rev.* *36*, 492–500.
- Branford, S., Yeung, D.T., Ross, D.M., Prime, J.A., Field, C.R., Altamura, H.K., Yeoman, A.L., Georgievski, J., Jamison, B.A., Phillis, S., et al. (2013). Early molecular response and female sex strongly predict stable undetectable BCR-ABL1, the criteria for imatinib discontinuation in patients with CML. *Blood* *121*, 3818–3824.
- Brooks, H.D., Glisson, B.S., Bekele, B.N., Johnson, F.M., Ginsberg, L.E., El-Naggar, A., Culotta, K.S., Takebe, N., Wright, J., Tran, H.T., et al. (2011). Phase 2 study of dasatinib in the treatment of head and neck squamous cell carcinoma. *Cancer* *117*, 2112–2119.
- Buggins, A.G., Mufti, G.J., Salisbury, J., Codd, J., Westwood, N., Arno, M., Fishlock, K., Pagliuca, A., and Devereux, S. (2002). Peripheral blood but not tissue dendritic cells express CD52 and are depleted by treatment with alemtuzumab. *Blood* *100*, 1715–1720.
- Carretero, J., Shimamura, T., Rikova, K., Jackson, A.L., Wilkerson, M.D., Borgman, C.L., Buttarazzi, M.S., Sanofsky, B.A., McNamara, K.L., Brandstetter, K.A., et al. (2010). Integrative genomic and proteomic analyses identify targets for Lkb1-deficient metastatic lung tumors. *Cancer Cell* *17*, 547–559.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* *2*, 401–404.
- Cook, M.B., McGlynn, K.A., Devesa, S.S., Freedman, N.D., and Anderson, W.F. (2011). Sex disparities in cancer mortality and survival. *Cancer Epidemiol. Biomarkers Prev.* *20*, 1629–1637.
- D’Agostino, R.B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat. Med.* *17*, 2265–2281.
- Dehm, S.M., and Bonham, K. (2004). SRC gene expression in human cancer: the role of transcriptional activation. *Biochem. Cell Biol.* *82*, 263–274.
- Deininger, M.W., O’Brien, S.G., Ford, J.M., and Druker, B.J. (2003). Practical management of patients with chronic myeloid leukemia receiving imatinib. *J. Clin. Oncol.* *21*, 1637–1647.
- Dimova, D.K., Stevaux, O., Frolov, M.V., and Dyson, N.J. (2003). Cell cycle-dependent and cell cycle-independent control of transcription by the *Drosophila* E2F/RB pathway. *Genes Dev.* *17*, 2308–2320.
- Dorak, T.M., and Karpuzoglu, E. (2012). Gender differences in cancer susceptibility: an inadequately addressed issue. *Front. Genet.* *3*, 268.
- Drolz, A., Wewalka, M., Horvatits, T., Fuhrmann, V., Schneeweiss, B., Trauner, M., and Zauner, C. (2014). Gender-specific differences in energy metabolism during the initial phase of critical illness. *Eur. J. Clin. Nutr.* *68*, 707–711.
- Fury, M.G., Baxi, S., Shen, R., Kelly, K.W., Lipson, B.L., Carlson, D., Stambuk, H., Haque, S., and Pfister, D.G. (2011). Phase II study of saracatinib (AZD0530) for patients with recurrent or metastatic head and neck squamous cell carcinoma (HNSCC). *Anticancer Res.* *31*, 249–253.
- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. *Nat. Rev. Cancer* *4*, 177–183.
- Ganesan, S., Silver, D.P., Greenberg, R.A., Avni, D., Drapkin, R., Miron, A., Mok, S.C., Randrianarison, V., Brodie, S., Salstrom, J., et al. (2002). BRCA1 supports XIST RNA concentration on the inactive X chromosome. *Cell* *111*, 393–405.
- Ho, D.E., Imai, K., King, G., and Stuart, E.A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Polit. Anal.* *15*, 199–236.
- Hsu, S.D., Lin, F.M., Wu, W.Y., Liang, C., Huang, W.C., Chan, W.L., Tsai, W.T., Chen, G.Z., Lee, C.J., Chiu, C.M., et al. (2011). miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.* *39*, D163–D169.
- Imbens, G.W. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review. *Rev. Econ. Stat.* *86*, 4–29.
- Janku, F., Wheler, J.J., Westin, S.N., Moulder, S.L., Naing, A., Tsimberidou, A.M., Fu, S., Falchook, G.S., Hong, D.S., Garrido-Laguna, I., et al. (2012). PI3K/AKT/mTOR inhibitors in patients with breast and gynecologic malignancies harboring PIK3CA mutations. *J. Clin. Oncol.* *30*, 777–782.
- Janku, F., Kaseb, A.O., Tsimberidou, A.M., Wolff, R.A., and Kurzrock, R. (2014). Identification of novel therapeutic targets in the PI3K/AKT/mTOR pathway in hepatocellular carcinoma using targeted next generation sequencing. *Oncotarget* *5*, 3012–3022.
- Jenne, D.E., Reimann, H., Nezu, J., Friedel, W., Loff, S., Jeschke, R., Muller, O., Back, W., and Zimmer, M. (1998). Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat. Genet.* *18*, 38–43.
- John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004). Human MicroRNA targets. *PLoS Biol.* *2*, e363.
- Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. *Cell* *115*, 787–798.
- Li, L., and Greene, T. (2013). A weighting analogue to pair matching in propensity score analysis. *Int. J. Biostat.* *9*, 215–234.
- Li, J., Lu, Y., Akbani, R., Ju, Z., Roebuck, P.L., Liu, W., Yang, J.Y., Broom, B.M., Verhaak, R.G., Kane, D.W., et al. (2013). TCGA: a resource for cancer functional proteomics data. *Nat. Methods* *10*, 1046–1047.
- Lipworth, L., Morgans, A.K., Edwards, T.L., Barocas, D.A., Chang, S.S., Herrell, S.D., Penson, D.F., Resnick, M.J., Smith, J.A., and Clark, P.E. (2014). Renal cell cancer histological subtype distribution differs by race and sex. *BJU Int.* *117*, 260–265.
- Liu, P., Cheng, H., Roberts, T.M., and Zhao, J.J. (2009). Targeting the phosphoinositide 3-kinase pathway in cancer. *Nat. Rev. Drug Discov.* *8*, 627–644.
- Mahoney, C.L., Choudhury, B., Davies, H., Edkins, S., Greenman, C., Haafren, G., Mironenko, T., Santarius, T., Stevens, C., Stratton, M.R., et al. (2009). LKB1/KRAS mutant lung cancers constitute a genetic subset of NSCLC with increased sensitivity to MAPK and mTOR signalling inhibition. *Br. J. Cancer* *100*, 370–375.
- Mao, J.H., Kim, I.J., Wu, D., Climent, J., Kang, H.C., DelRosario, R., and Balmain, A. (2008). FBXW7 targets mTOR for degradation and cooperates with PTEN in tumor suppression. *Science* *321*, 1499–1502.
- Marchetti, A., Martella, C., Felicioni, L., Barassi, F., Salvatore, S., Chella, A., Camplese, P.P., Iarussi, T., Mucilli, F., Mezzetti, A., et al. (2005). EGFR

- mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. *J. Clin. Oncol.* **23**, 857–865.
- Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhi, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41.
- Mittelstrass, K., Ried, J.S., Yu, Z., Krumsiek, J., Gieger, C., Prehn, C., Roemisch-Margl, W., Polonikov, A., Peters, A., Theis, F.J., et al. (2011). Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet.* **7**, e1002215.
- Molife, R., Lorigan, P., and MacNeil, S. (2001). Gender and survival in malignant tumours. *Cancer Treat. Rev.* **27**, 201–209.
- Nakayama, S., Sng, N., Carretero, J., Welner, R., Hayashi, Y., Yamamoto, M., Tan, A.J., Yamaguchi, N., Yasuda, H., Li, D., et al. (2014).  $\beta$ -catenin contributes to lung tumor development induced by EGFR mutations. *Cancer Res.* **74**, 5891–5902.
- Nissan, M.H., Pratilas, C.A., Jones, A.M., Ramirez, R., Won, H., Liu, C., Tiwari, S., Kong, L., Hanrahan, A.J., Yao, Z., et al. (2014). Loss of NF1 in cutaneous melanoma is associated with RAS activation and MEK dependence. *Cancer Res.* **74**, 2340–2350.
- Ostrom, Q.T., Gittleman, H., Liao, P., Rouse, C., Chen, Y., Dowling, J., Wolinsky, Y., Kruchko, C., and Barnholtz-Sloan, J. (2014). CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007–2011. *Neuro Oncol.* **16** (Suppl 4), iv1–iv63.
- Pal, S.K., and Hurria, A. (2010). Impact of age, sex, and comorbidity on cancer therapy and disease progression. *J. Clin. Oncol.* **28**, 4086–4093.
- Pardoll, D.M. (2012). The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* **12**, 252–264.
- Purtilo, D.T., and Sullivan, J.L. (1979). Immunological bases for superior survival of females. *Am. J. Dis. Child.* **133**, 1251–1253.
- Rosenbaum, P.R., and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.
- Schuetz, W., Schirmacher, P., Eberhardt, W.E., Fischer, J.R., von der Schulenburg, J.M., Mezger, J., Schumann, C., Serke, M., Zaun, S., Dietel, M., et al. (2015). EGFR mutation status and first-line treatment in patients with stage III/IV non-small cell lung cancer in Germany: an observational study. *Cancer Epidemiol. Biomarkers Prev.* **24**, 1254–1261.
- Shackelford, D.B., Abt, E., Gerken, L., Vasquez, D.S., Seki, A., Leblanc, M., Wei, L., Fishbein, M.C., Czernin, J., Mischel, P.S., et al. (2013). LKB1 inactivation dictates therapeutic response of non-small cell lung cancer to the metabolism drug phenformin. *Cancer Cell* **23**, 143–158.
- Shepherd, F.A., Pereira, J.R., Ciuleanu, T., Tan, E.H., Hirsh, V., Thongprasert, S., Campos, D., Maoleekoonpiroj, S., Smylie, M., Martins, R., et al. (2005). Erlotinib in previously treated non-small-cell lung cancer. *N. Engl. J. Med.* **353**, 123–132.
- Siegel, R.L., Miller, K.D., and Jemal, A. (2015). Cancer statistics, 2015. *CA Cancer J. Clin.* **65**, 5–29.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550.
- Suwaki, N., Vanhecke, E., Atkins, K.M., Graf, M., Swabey, K., Huang, P., Schraml, P., Moch, H., Cassidy, A.M., Brewer, D., et al. (2011). A HIF-regulated VHL-PTP1B-Src signaling axis identifies a therapeutic target in renal cell carcinoma. *Sci. Translational Med.* **3**, 85ra47.
- Tam, I.Y.S., Chung, L.P., Suen, W.S., Wang, E., Wong, M.C.M., Ho, K.K., Lam, W.K., Chiu, S.W., Girard, L., Minna, J.D., et al. (2006). Distinct epidermal growth factor receptor and KRAS mutation patterns in non-small cell lung cancer patients with different tobacco exposure and clinicopathologic features. *Clin. Cancer Res.* **12**, 1647–1653.
- Tenbaum, S.P., Ordonez-Moran, P., Puig, I., Chicote, I., Arques, O., Landolfi, S., Fernandez, Y., Herance, J.R., Gispert, J.D., Mendizabal, L., et al. (2012).  $\beta$ -catenin confers resistance to PI3K and AKT inhibitors and subverts FOXO3a to promote metastasis in colon cancer. *Nat. Med.* **18**, 892–901.
- Tennyson, C.N., Klamut, H.J., and Worton, R.G. (1995). The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat. Genet.* **9**, 184–190.
- The Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120.
- The Cancer Genome Atlas Research Network. (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550.
- Valeyrie, L., Bastuji-Garin, S., Revuz, J., Bachot, N., Wechsler, J., Berthaud, P., Tulliez, M., and Giraudier, S. (2003). Adverse cutaneous reactions to imatinib (STI571) in Philadelphia chromosome-positive leukemias: a prospective study of 54 patients. *J. Am. Acad. Dermatol.* **48**, 201–206.
- Van Allen, E.M., Wagle, N., Stojanov, P., Perrin, D.L., Cibulskis, K., Marlow, S., Jane-Valbuena, J., Friedrich, D.C., Kryukov, G., Carter, S.L., et al. (2014). Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat. Med.* **20**, 682–688.
- Van der Meulen, J., Sanghvi, V., Mavrakis, K., Durinck, K., Fang, F., Matthijssens, F., Rondou, P., Rosen, M., Pieters, T., Vandenberghe, P., et al. (2015). The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. *Blood* **125**, 13–21.
- Vincent-Salomon, A., Ganem-Elbaz, C., Manie, E., Raynal, V., Sastre-Garau, X., Stoppa-Lyonnet, D., Stern, M.H., and Heard, E. (2007). X inactive-specific transcript RNA coating and genetic instability of the X chromosome in BRCA1 breast tumors. *Cancer Res.* **67**, 5134–5140.
- Weinstein, Y., Ran, S., and Segal, S. (1984). Sex-associated differences in the regulation of immune responses controlled by the MHC of the mouse. *J. Immunol.* **132**, 656–661.
- Wertz, I.E., Kusam, S., Lam, C., Okamoto, T., Sandoval, W., Anderson, D.J., Helgason, E., Ernst, J.A., Eby, M., Liu, J., et al. (2011). Sensitivity to antitubulin chemotherapeutics is regulated by MCL1 and FBW7. *Nature* **471**, 110–114.
- Wong, N., and Wang, X. (2015). miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* **43**, D146–D152.
- Younes, A., Connors, J.M., Park, S.I., Fanale, M., O'Meara, M.M., Hunder, N.N., Huebner, D., and Ansell, S.M. (2013). Brentuximab vedotin combined with ABVD or AVD for patients with newly diagnosed Hodgkin's lymphoma: a phase 1, open-label, dose-escalation study. *Lancet Oncol.* **14**, 1348–1356.
- Zhang, S., Liu, X., Bawa-Khalife, T., Lu, L.S., Lyu, Y.L., Liu, L.F., and Yeh, E.T. (2012). Identification of the molecular basis of doxorubicin-induced cardiotoxicity. *Nat. Med.* **18**, 1639–1642.