Cancer initiation is influenced by sex-biased tissue environments and hormonal imbalances

Feng Zhang[1], Minghua Deng[1,2,3]*

Affiliations:

[1]Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China.
[2]School of Mathematical Sciences, Peking University, Beijing, China.
[3]Center for Statistical Science, Peking University, Beijing, China.
*Corresponding author. Email: dengmh@math.pku.edu.cn.

Abstract

There is extensive evidence of sex differences in the susceptibility and prognosis of non-reproductive cancers. In addition to external factors, biological sex bias (e.g., sex chromosomes, hormones, and immune function) is suspected to function as a selective pressure that influences cancer initiation. However, there remains a lack of clarity concerning the extent of the effect of sex bias on cancer initiation, as well as the underlying mechanism. In this study, we show that tissue sex bias—directed by gonadotropin-releasing hormone—varies among tissues and is associated with two distinct age-specific patterns of cancer incidence: parallel and nonparallel. Additionally, we reveal that imbalances in estrogen and thyroid hormone are correlated with levels of hypoxia-inducible factors, which have three phases (hypoxia, hyperoxia, and "chaotic-oxia") that exist in most cancers and are linked to specific cancer subtypes, including cancers with microsatellites, the CpG island methylator phenotype, or hypermethylation. Our results suggest that sex-biased tissue environments and hormonal imbalances influence the evolutionary dynamics of cancer initiation, emphasizing the importance of maintaining hormonal homeostasis for cancer prevention and providing insights toward improving therapies for cancer types that involve hormonal imbalances.

Main

The age-specific incidence curve of a particular cancer can represent the evolutionary dynamics of cancer initiation. For carcinomas (i.e., most solid cancers), age-specific incidence curves are approximately linear on a log-log scale *(1)*. Armitage and Doll *(2)* proposed a multistage model explaining the power law relationship, based on the assumption that carcinogenesis is a multistage process involving the accumulation of driver mutations. This model has been modified to incorporate new information generated by other studies *(3-8)*. Cancer initiation is influenced by selective pressure from the microenvironment, through a process in which driver mutations (genetic or epigenetic) are selected and accumulate. Generally, different cancer types have distinct mutational spectra *(9,10)*, indicating that variations in tissue environment and cell type can influence the evolutionary dynamics of cancer initiation.

For nearly all non-reproductive cancers, women and men have different susceptibilities and prognoses. However, behavior or mutagen exposure can only partly explain this phenomenon *(11)*. Biological sex bias, involving sex chromosomes and hormones, also plays a role in many instances *(12-15)*. Recent studies have shown that female and male patients preferentially exhibit mutations of certain genes and gain specific transcriptional or epigenetic signatures *(16-18)*; because they display stronger immunity, female patients preferentially accumulate more mutant peptides that are not recognized by T cell receptors *(19)*. These studies provide evidence to support the hypothesis that sex bias is a selective pressure with the potential to influence cancer initiation. However, the mechanism by which sex bias influences cancer initiation remains largely unknown.

Parallel and nonparallel patterns of age-specific cancer incidence in women and men

Based on the hypothesis that sex bias influences cancer initiation which can be represented by the age-specific incidence curve of a particular cancer, we used the CI5-X database (provided by the International Agency for Research on Cancer) to compare the incidence curves of major solid cancer types between women and men in selected regions or ethnicities (see the supplementary materials). Fitted age-specific incidence curves for men and women showed two general patterns on a log-log scale: parallel and nonparallel (Fig. 1A). Parallel patterns were observed when colon cancer incidences were compared between inherited and sporadic types *(7)*, and when lung cancer incidences were compared between smokers and non-smokers *(20)*. These parallel patterns suggest that individuals with a susceptible genetic background or greater exposure to mutagens have a higher baseline risk of cancer. To our knowledge, nonparallel patterns have not been reported.

To determine the prevalence of nonparallel patterns in major solid cancers according to region or ethnicity, we conducted a linear regression analysis and explored age-by-sex interactions to determine the statistical significance of the nonparallel pattern in each cohort (see the supplementary materials). Our analysis revealed that nonparallel patterns were more common in stomach cancer, melanoma, and thyroid cancer (Fig. 1B).

Sex bias in tissue environments

A potential explanation for the prevalence of nonparallel patterns is that tissue environments in the stomach, skin, and thyroid exhibit greater sex differences compared with tissue environments in the other regions; these differences may influence cancer initiation. To test this hypothesis, we used GTEx data to evaluate sex bias in tissue environments via sex-biased expressed pathway analysis. In our analysis, the level of expression in a pathway was represented by the mean level

of expression for all genes in that pathway; we randomly selected 100 samples from each tissue to maintain consistent statistical power across tissues. We explored sex-biased expressed pathway using limma *(21)*; we repeated this process 1000 times and recorded the numbers of sex-biased expressed pathways in each tissue during each repetition (see the supplementary materials). There were fewer than 100 kidney and bladder samples (Table S1) and no significant sex-biased expressed pathways were identified in those tissues; thus, their results were regarded as not available (NA). The results showed that the adrenal glands, stomach, thyroid, and skin contained the largest numbers of sex-biased expressed pathways (Fig. 2A).

To determine which pathways have greater involvement in tissue sex bias, we re-examined sex-biased expressed pathways using all samples in each type of tissue except the bladder, for which only nine samples were available. Among the pathways, only the gonadotropin-releasing hormone (GnRH) secretion pathway (hsa04929) exhibited significant sex bias (limma $P < 0.05$) in the top three tissues with sex bias (adrenal glands, thyroid, and stomach), but not in other tissues (P for skin = 0.0683, Table S2). Additionally, the female-to-male log fold changes (LogFCs) of the GnRH secretion pathway were highest in the adrenal glands, stomach, and thyroid; the LogFC in skin was fifth highest (Fig. 2B). The LogFC of the GnRH secretion pathway was positively correlated (Pearson correlation coefficient $R = 0.74$) with the number of sex-biased expressed pathways in each tissue. The LogFC of another GnRH-related pathway, the GnRH signaling pathway (hsa04912), was also positively correlated (Pearson correlation coefficient $R = 0.91$) with the number of sex-biased expressed pathways in each tissue. The adrenal glands, stomach, and thyroid had the highest LogFCs, whereas the skin had a negative LogFC (Fig. 2C, Table S3).

Adrenal gland cancer is a rare type of cancer, and linear representation of its age-specific incidence is hindered by noise. Therefore, it was not included in the nonparallel analysis shown in Figure 1B. However, because adrenal gland tissue had the highest number of sex-biased expressed pathways and largest LogFC of the two GnRH-related pathways, we repeated the nonparallel pattern analysis with a focus on adrenal gland cancer. To reduce noise in the incidence of adrenal gland cancer, we merged the regional cohorts into continental cohorts (see the supplementary materials). We found that adrenal cancer incidence also displayed a nonparallel pattern; the results were statistically significant in North American and Asian cohorts (Fig. 2D).

The most prevalent nonparallel patterns in stomach cancer, melanoma, thyroid cancer, and adrenal cancer were observed across regions or ethnicities (Figs. 1B and 2D). Consistent with these findings, adrenal glands, stomach, thyroid, and skin exhibited the highest numbers of sex-biased expressed pathways (Fig. 2A), which were associated with GnRH-related pathways. Thus, sex-biased tissue environments may influence the evolutionary dynamics of cancer initiation and induce nonparallel patterns.

Genes with sex-biased mutation frequency

Because cancer initiation can be influenced by sex, there may exist genes that exhibit a greater likelihood of mutation in a single sex (hereafter, “sex-biased mutated genes”). Previous studies attempted to explore this hypothesis, but they had less statistical power because of reliance on incomplete The Cancer Genome Atlas (TCGA) data *(16, 17)* or whole genome sequencing data alone *(18)*. Therefore, we used complete TCGA data to conduct a comparative analysis of gene mutation frequency between women and men among 6181 individuals across 24 non-reproductive cancers; we utilized both the likelihood ratio test (LRT) and logistic regression

analysis to adjust for confounding factors (see the supplementary materials). Our results revealed eight sex-biased mutated genes (Table 1): DNAH7, LARP4B, PGM5, and PTEN in stomach adenocarcinoma (STAD); SELP and SPAG17 in skin cutaneous melanoma (SKCM); RBM10 in lung adenocarcinoma (LUAD); and CASP8 in head and neck squamous cell carcinoma (HNSC).

Hotspot dominance in three sex-biased mutated genes in STAD

In STAD, three sex-biased mutated genes (PTEN, LARP4B, and PGM5) displayed hotspot mutation dominance, as shown in Figure 3 (depicted using ProteinPaint) *(22)*. These hotspots, Lys267ArgfsTer9 (c.800delA) in PTEN, Thr163HisfsTer47 (c.487delA) in LARP4B, and Ile98Val (c.292A>G) in PGM5, also showed a significant sex bias ($P < 0.05$, Fisher's exact test, Fig. 3). TCGA data showed that these three hotspots were exclusively present in STAD, colon adenocarcinoma (COAD), and uterine corpus endometrial carcinoma (UCEC), with the exception of a PGM5 hotpot in prostate adenocarcinoma (PRAD). The mutational spectra of other sex-biased mutated genes are shown in Figure S1.

Sex-biased hotspots and sex-biased proportions of cancers with microsatellite instability (MSI)

To determine the underlying cause of sex bias in the presence of the three hotspots, we conducted co-occurrence and mutual exclusivity analyses of the three hotspots in STAD, COAD, and UCEC; the results suggested that the three sex-biased hotspots tend to co-occur (Table S4). To gain further insight, we divided the samples into two groups (HOTSPOT+ and HOTSPOT-) according to the presence or absence of at least one of the three hotspots across STAD, COAD, and UCEC. Subsequently, we used limma to conduct an analysis of differentially expressed genes (DEGs) between the two groups.

In all three cancers, four genes were differentially expressed between the two groups: C7orf31, MLH1, EPM2AIP1, and LINC00667 (Data S4). MLH1, a well-known tumor suppressor gene, showed reduced expression in the HOTSPOT+ group across all three cancers (Fig. S2). MLH1 deficiency can induce MSI, which is commonly observed in colorectal, gastric, and endometrial cancer. The signatures of the three hotspots were consistent with previously reported mutational signatures of MSI cancer types: the PTEN and LARP4B hotspots are typical of the MSI mutational signature ID1, whereas the PGM5 hotspot is typical of the MSI mutational signatures SBS21 and SBS26 *(23)*. Furthermore, all HOTSPOT+ samples were labeled as the MSI subtype in TCGA data (see the supplementary materials). This evidence suggests that the three hotspots result from hypermutation across MSI subtypes.

We speculated that hotspot-specific sex bias is caused by sex bias in MSI subtype. Thus, we examined MSI subtype proportions in STAD and COAD; in both cancers, we found that the MSI subtype proportion was higher in women than in men. The difference was statistically significant in STAD (Table S5). A previous study *(24)* revealed that, in both gastric and colorectal cancers, the MSI subtype proportion was significantly higher in women than in men (Table S6).

Estrogen receptor alpha (ESR1), thyroid hormone receptors (THRs), and hypoxia-inducible factors (HIFs) in cancers with MSI

The above findings led us to investigate the underlying cause of sex bias in MSI subtype. For this purpose, we divided STAD, COAD, and UCEC samples into two groups: MSI-positive (MSI+) and MSI-negative (MSI-). We then analyzed DEGs in each cancer type (see the supplementary materials).

Hypoxia is known to suppress mismatch repair through mechanisms such as the reduction of MLH1 expression *(25-28)*. In STAD, COAD, and UCEC, we observed that MSI+ samples consistently had a significantly elevated level of HIF1A (a key activator of hypoxia signaling) or a reduced level of HIF3A (a repressor of hypoxia signaling) *(29)*, compared with MSI- samples. These findings indicated that MSI+ samples had a higher level of hypoxia. Subsequently, we used the ratio of HIF1A to HIF3A (i.e., HIF level), to represent the level of hypoxia in each sample. We found that the HIF level was significantly higher in MSI+ samples than in MSI- samples; it was positively correlated with the MLH1 methylation level in all three cancers (Fig. 4A).

ESR1 has been reported that it can directly regulates HIF1A, indicating crosstalk between the hypoxia and estrogen signaling pathways *(30)*. However, in STAD, COAD, and UCEC, we found that the level of ESR1 did not significantly differ between MSI+ and MSI- samples. Furthermore, the expression of ESR1 was positively correlated with the HIF level in UCEC, but not in STAD or COAD (Figs. 4B and 5A). Finally, pan-cancer analysis of TCGA data showed that ESR1 only had a slight positive correlation with the HIF level; conversely, pan-tissue analysis of GTEx data revealed a negative correlation with the HIF level (Fig. S3).

Our earlier results indicated that GnRH was involved in the establishment of a sex-biased tissue environment. GnRH is a key regulator of the hypothalamic-pituitary-gonadal (HPG) axis, which is upstream of estrogen signaling and has substantial crosstalk with the hypothalamic-pituitary-adrenal (HPA) and hypothalamic-pituitary-thyroid (HPT) axes *(31)*. Considering these connections, besides ESR1, we suspected that other hormones or hormone receptors were involved in the sex bias in MSI subtype. Therefore, we investigated the expression patterns of a set of genes involved in the HPG, HPA, and HPT axes (see the supplementary materials).

The results showed that only one gene, thyroid hormone receptor beta (THRB), was significantly differentially expressed between MSI+ and MSI- samples in all three cancer types. Moreover, thyroid hormone receptor alpha (THRA) was significantly differentially expressed in STAD and COAD, but not in UCEC (Fig. S4, Data S5). Because THRA and THRB are both important receptors for thyroid hormone (TH), we used the summed logarithmic levels of THRA and THRB to represent the THR level in each sample. We found that the THR level was significantly lower in MSI+ samples than in MSI- samples; it was negatively correlated with the HIF level in all three cancers (Fig. 4C). The negative correlation between HIF and THR was also confirmed by pan-cancer analysis of TCGA data and pan-tissue analysis of GTEx data (Fig. 4D).

ESR1, THR, and HIF in cancers with hypermethylation or in estrogen receptor-positive (ER+) breast cancers (BRCAs)

Epigenetic silencing of the MLH1 gene is a major cause of MSI; it can occur in cancers with the CpG island methylator phenotype (CIMP) or hypermethylation. There is a significant overlap between colorectal cancers with MSI and colorectal cancers with CIMP *(32)*. Accordingly, to explore whether cancers with CIMP or hypermethylation exhibit expression signatures similar to MSI+ cancers (e.g., ESR1, THR, and HIF), we conducted a similar analysis of cancers with sufficient CIMP or hypermethylation labels in the TCGA dataset. These cancers were designated “HyperMeth cancers,” whereas other cancers were designated “HypoMeth cancers” (see the supplementary materials).

Based on their expression signatures, HyperMeth cancers were divided into two subgroups: hypoxia and hyperoxia. In the hypoxia subgroup, similar to MSI+ cancers, all HyperMeth cancers demonstrated a significantly elevated HIF level (except HyperMeth SKCM, in which the

increase was not statistically significant); they typically had a reduced THR level or an elevated ESR1 level (Figs. 5A and S6). In the hyperoxia subgroup, all HyperMeth cancers demonstrated a significantly reduced HIF level (except HyperMeth THCA, in which the decrease was not statistically significant); they also had a significantly elevated THR level or a significantly reduced ESR1 level (Fig. 5A).

Adrenocortical carcinoma (ACC) and pheochromocytoma and paraganglioma (PCPG) both originate from adrenal glands. However, HyperMeth ACC and HyperMeth PCPG displayed opposing THR and HIF expression signatures. Specifically, HyperMeth ACC exhibited hypoxia, whereas HyperMeth PCPG exhibited hyperoxia. This difference suggests that the two cancers originated under divergent selective pressures or evolved in opposite directions.

BRCA is classified into four well-characterized molecular subtypes: luminal A, luminal B, HER2-positive, and triple-negative. Estrogen plays a crucial role in the development of ER+ BRCAs (i.e., luminal A and luminal B). To investigate whether ER+ BRCAs exhibit hypoxia and have expression signatures similar to MSI+ or hypoxic HyperMeth cancers, we compared expression signatures between ER+ BRCAs and ER- (i.e., HER2-positive and triple-negative) BRCAs. In contrast to the findings in MSI+ and HyperMeth cancers, we observed an antagonistic expression pattern in ER+ BRCAs: ESR1 and THR levels were both significantly elevated, whereas the HIF level did not significantly differ (Fig. 5A). However, the HIF1A level tended to be significantly reduced, and the HIF3A level was significantly reduced (Fig. S5).

Estrogen receptor beta (ESR2) is also involved in hypoxia signaling. However, its role is not well established, and it exhibits opposing roles in some cancers *(33)*. Our results showed that ESR2 was significantly elevated in MSI+ COAD and significantly reduced in HyperMeth SKCM, indicating that ESR2 may play a role in the initiation of these cancers.

Additionally, similar to the three hotspot mutations exclusively found in MSI+ cancers, some other sex-biased mutated genes tend to be present in certain cancer subtypes. Specifically, CASP8 was typically mutated in HyperMeth HNSC, SELP was typically mutated in HyperMeth SKCM, and DNAH7 was typically mutated in MSI+ STAD. These results were statistically significant (Table S7), suggesting a link between these sex-biased mutated genes and the corresponding cancer subtypes.

Imbalances in estrogen and TH may contribute to carcinogenesis

The above results demonstrated that a sex-biased tissue environment can influence cancer initiation. ESR1 and THR are associated with the HIF level in cancer, as well as specific cancer subtypes, suggesting that they also play a role in cancer initiation. ESR1 can directly regulate HIF1A and induce the hypoxia signaling pathway *(30)*. TH has a more complex role in hypoxia. TH can induce HIF1A activity, which then reduces TH signaling; however, experimental and clinical studies have revealed that TH plays a critical role in protecting tissues from hypoxic injury *(34)*. We also observed this complex phenomenon: analysis of GTEx data revealed that the THR level was negatively correlated with the HIF1A level but positively correlated with the HIF3A level (Fig. S7). However, analysis of TCGA data demonstrated that the negative correlation between THR and HIF1A disappeared, whereas the positive correlation between THR and HIF3A persisted (Fig. S8). These results indicated that THR generally functions as a hypoxia repressor in cancers, rather than a hypoxia activator, similar to its protection against hypoxic injury.

ERs and THRs are both nuclear receptors with similar DNA-binding sequences. They can competitively bind to each other's DNA-binding sites and inhibit subsequent transcription *(35)*. The mutual antagonism of ERs and THRs is consistent with the observations in Figure 5A: an increase in ESR1 or a decrease in THR is accompanied by hypoxia, whereas a decrease in ESR1 or an increase in THR is accompanied by hyperoxia.

Based on the present findings, we propose a model of carcinogenesis that involves imbalances in estrogen and TH (Fig. 5B). As noted above, TH and estrogen play key roles in many aspects of biological processes and maintain homeostasis under healthy conditions. However, TH or estrogen dominance can induce changes in the tissue environment, such as hypoxia or hyperoxia; it can also alter the epigenome and transcriptome in tissue cells. In ER+ BRCAs, both TH and estrogen can become simultaneously elevated, resulting in a state of antagonism or "chaotic-oxia." In terms of selective pressure for cancer evolution, all of these altered tissue environments can induce specific cancer subtypes, as shown in Figure 5A and B.

Discussion

The present study showed that cancer initiation can be influenced by sex bias in two ways: sex-biased tissue environments and hormonal imbalances. However, these influences are not mutually exclusive and may interact with each other. Here, we found that GnRH was associated with a sex-biased tissue environment; GnRH is a key regulator of the HPG axis and receives negative feedback from gonadal steroids *(36)*. The HPG axis has extensive crosstalk with the HPT axis, such as crosstalk between the estrogen and TH signaling pathways *(31, 35)*. Together, the HPG axis, HPT axis, and other factors regulate tissue environment homeostasis. Generally, the level of estrogen is higher in women than in men; this increases the likelihood of a hypoxic tissue environment in women and the potential for cancers with MSI. As expected, we found that women had a higher prevalence of MSI in STAD and COAD; we also found that women had a high prevalence of MSI in UCEC, where estrogen is suspected to promote cancer progression (Table S5 and S6). In contrast, the level of free triiodothyronine is generally higher in men than in women; this increases the likelihood of a hyperoxic tissue environment in men and the potential for cancers such as the hyperoxic CIMP subtype of thyroid carcinoma. Although the difference was not statistically significant (possibly because of the small sample size), we found that men had a higher prevalence of the hyperoxic CIMP subtype of thyroid carcinoma (Table S5). However, we found that the hypoxic HyperMeth subtype of HNSC was more prevalent in men, indicating that it may exhibit tissue-specific regulation. These results suggest that in the context of sex-biased tissue environments and hormonal imbalances, cancer precursors have a preferred evolutionary trajectory. Consistent with this hypothesis, we observed that six of the eight sex-biased mutated genes were significantly enriched in specific cancer subtypes. However, in cancers such as PCPG and SKCM, the proportions of subtypes were similar in women and men.

In addition to influencing the direction of cancer evolution, sex-biased tissue environments and hormonal imbalances may also affect the dynamics of cancer evolution. GnRH secretion exhibits pulsatile behavior in both men and women, but the frequency and amplitude of pulses in women are dependent on the phase of the menstrual cycle. Moreover, women display a unique GnRH secretion pattern, which involves a surge during the pre-ovulatory phase *(36)*. The sex-biased dynamics of GnRH secretion may lead to sex-biased dynamics of cancer evolution. Hypermutation or hypermethylation, which occur in cancers with hormonal imbalances, could lead to more rapid cancer evolution and altered dynamics, as described in a previous study *(6)*. All of these factors may contribute to the nonparallel pattern of cancer incidence observed in this

study. However, greater understanding of sex-biased biological mechanisms is needed to develop a detailed mathematical model of sex bias patterns.

TH is suspected to play a role in cancer initiation. For example, compared with men, women are 10-fold more likely to develop hyperthyroidism and hypothyroidism; these disorders have been linked to changes in the incidence and prognosis of many cancers *(37, 38)*. The present findings support the increasing importance of TH in cancer research. Although our study focused on the balance between TH and estrogen, we acknowledge that other hormones may also contribute to cancer initiation. For example, progestogen is reportedly involved in luminal A BRCA, and androgens have been linked to prostate cancer.

In this study, samples were grouped by molecular clustering labels, which may not fully capture the status of the balance between estrogen and TH. The identification of molecular patterns associated with imbalances in estrogen and TH could enhance our findings and reveal additional mechanisms involving sex bias. Furthermore, the levels of ERs and THRs may not directly reflect the levels of estrogen and TH. Nevertheless, this study showed that three phases (hypoxia, hyperoxia, and chaotic-oxia) are present in most cancers and are linked to hormonal imbalances. These findings emphasize the importance of maintaining hormonal homeostasis for cancer prevention and provide insights toward improving therapies for cancer types that involve hormonal imbalances.

References and Notes

1 Nordling, C. O. A New Theory on the Cancer-Inducing Mechanism. *British Journal of Cancer* **7**, 68-72 (1953).

2 Armitage, P. & Doll, R. The Age Distribution of Cancer and a Multi-Stage Theory of Carcinogenesis. *British Journal of Cancer* **8**, 1-12 (1954).

3 Knudson, A. G. Mutation and Cancer - Statistical Study of Retinoblastoma. *P Natl Acad Sci USA* **68**, 820-& (1971).

4 Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23-28, doi:10.1126/science.959840 (1976).

5 Fearon, E. R. & Vogelstein, B. A Genetic Model for Colorectal Tumorigenesis. *Cell* **61**, 759-767 (1990).

6 Michor, F., Iwasa, Y. & Nowak, M. A. Dynamics of cancer progression. *Nat Rev Cancer* **4**, 197-205, doi:10.1038/nrc1295 (2004).

7 Frank, S. A. Age-specific incidence of inherited versus sporadic cancers: A test of the multistage theory of carcinogenesis. *P Natl Acad Sci USA* **102**, 1071-1075 (2005).

8 Meza, R., Jeon, J., Moolgavkar, S. H. & Luebeck, E. G. Age-specific incidence of cancer: Phases, transitions, and biological implications. *P Natl Acad Sci USA* **105**, 16284-16289, doi:10.1073/pnas.0801151105 (2008).

9 Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339**, 1546-1558 (2013).

10 Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science* **349**, 1483-1489 (2015).

11 Jackson, S. S. *et al.* Sex disparities in the incidence of 21 cancer types: Quantification of the contribution of risk factors. *Cancer-Am Cancer Soc* **128**, 3531-3540, doi:10.1002/cncr.34390 (2022).

12 Clocchiatti, A., Cora, E., Zhang, Y. & Dotto, G. P. Sexual dimorphism in cancer. *Nat Rev Cancer* **16**, 330-339 (2016).

13 Rubin, J. B. *et al.* Sex differences in cancer mechanisms. *Biol Sex Differ* **11**, doi:10.1186/s13293-020-00291-x (2020).

14 Lopes-Ramos, C. M., Quackenbush, J. & DeMeo, D. L. Genome-Wide Sex and Gender Differences in Cancer. *Front Oncol* **10**, doi:10.3389/fonc.2020.597788 (2020).

15 Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369**, 1331-+, doi:10.1126/science.aba3066 (2020).

16 Yuan, Y. *et al.* Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients. *Cancer Cell* **29**, 711-722, doi:10.1016/j.ccell.2016.04.001 (2016).

17 Dunford, A. *et al.* Tumor-suppressor genes that escape from X-inactivation contribute to cancer sex bias. *Nature Genetics* **49**, 10-16, doi:10.1038/ng.3726 (2017).

18 Li, C. H. *et al.* Sex differences in oncogenic mutational processes. *Nat Commun* **11**, 4330, doi:10.1038/s41467-020-17359-2 (2020).

19 Castro, A. *et al.* Strength of immune selection in tumors varies with sex and age. *Nature Communications* **11** (2020).

20 Doll, R. The Age Distribution of Cancer: Implications for Models of Carcinogenesis. *Journal of the Royal Statistical Society. Series A (General)* **134**, 133, doi:10.2307/2343871 (1971).

21 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).

22 Zhou, X. *et al.* Exploring genomic alterations in pediatric cancer using ProteinPaint. *Cancer Res* **76** (2016).

23 Mas-Ponte, D., McCullough, M. & Supek, F. Spectrum of DNA mismatch repair failures viewed through the lens of cancer genomics and implications for therapy. *Clin Sci (Lond)* **136**, 383-404, doi:10.1042/CS20210682 (2022).

24 Akagi, K. *et al.* Real-world data on microsatellite instability status in various unresectable or metastatic solid tumors. *Cancer Science* **112**, 1105-1113, doi:10.1111/cas.14798 (2021).

25 Kaplan, A. R. & Glazer, P. M. Impact of hypoxia on DNA repair and genome integrity. *Mutagenesis* **35**, 61-67, doi:10.1093/mutage/gez019 (2020).

26 Mihaylova, V. T. *et al.* Decreased expression of the DNA mismatch repair gene Mlh1 under hypoxic stress in mammalian cells. *Mol Cell Biol* **23**, 3265-3273, doi:Doi 10.1128/Mcb.23.9.3265-3273.2003 (2003).

27 Lu, Y., Wajapeyee, N., Mitchell & Peter. Silencing of the DNA Mismatch Repair Gene MLH1 Induced by Hypoxic Stress in a Pathway Dependent on the Histone Demethylase LSD1. *Cell Reports* **8**, 501-513, doi:10.1016/j.celrep.2014.06.035 (2014).

28 Chen, H. B., Yan, Y., Davidson, T. L., Shinkai, Y. & Costa, M. Hypoxic stress induces dimethylated histone H3 lysine 9 through histone methyltransferase G9a in mammalian cells. *Cancer Res* **66**, 9009-9016, doi:10.1158/0008-5472.Can-06-0101 (2006).

29 Duan, C. Hypoxia-inducible factor 3 biology: complexities and emerging themes. *Am J Physiol Cell Physiol* **310**, C260-269, doi:10.1152/ajpcell.00315.2015 (2016).

30 Yang, J., Harris, A. L. & Davidoff, A. M. Hypoxia and Hormone-Mediated Pathways Converge at the Histone Demethylase KDM4B in Cancer. *International Journal of Molecular Sciences* **19**, doi:10.3390/ijms19010240 (2018).

31 Thambirajah, A. A. *et al.* Disruption by stealth-Interference of endocrine disrupting chemicals on hormonal crosstalk with thyroid axis function in humans and other animals. *Environ Res* **203**, doi:10.1016/j.envres.2021.111906 (2022).

32 Ogino, S. & Goel, A. Molecular classification and correlates in colorectal cancer. *J Mol Diagn* **10**, 13-27, doi:10.2353/jmoldx.2008.070082 (2008).

33 Mal, R. *et al.* Estrogen Receptor Beta (ER beta): A Ligand Activated Tumor Suppressor. *Front Oncol* **10**, doi:10.3389/fonc.2020.587386 (2020).

34 Lourbopoulos, A. I., Mourouzis, I. S., Trikas, A. G., Tseti, I. K. & Pantos, C. I. Effects of Thyroid Hormone on Tissue Hypoxia: Relevance to Sepsis Therapy. *J Clin Med* **10**, doi:10.3390/jcm10245855 (2021).

35 Ren, B. & Zhu, Y. A New Perspective on Thyroid Hormones: Crosstalk with Reproductive Hormones in Females. *Int J Mol Sci* **23**, doi:10.3390/ijms23052708 (2022).

36 Tsutsumi, R. & Webster, N. J. G. GnRH Pulsatility, the Pituitary Response and Reproductive Dysfunction. *Endocr J* **56**, 729-737, doi:DOI 10.1507/endocrj.K09E-185 (2009).

37 Taylor, P. N. *et al.* Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol* **14**, 301-316, doi:10.1038/nrendo.2018.18 (2018).

38 Krashin, E., Piekielko-Witkowska, A., Ellis, M. & Ashur-Fabian, O. Thyroid Hormones and Cancer: A Comprehensive Review of Preclinical and Clinical Studies. *Front Endocrinol (Lausanne)* **10**, 59, doi:10.3389/fendo.2019.00059 (2019).

39 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *Ca-a Cancer Journal for Clinicians* **69**, 7-34 (2019).

40 Nunney, L. Commentary: The multistage model of carcinogenesis, Peto's paradox and evolution. *International Journal of Epidemiology* **45**, 649-U667, doi:10.1093/ije/dyv201 (2016).

41 Moolgavkar, S. H. Commentary: Fifty years of the multistage model: remarks on a landmark paper. *Int J Epidemiol* **33**, 1182-1183, doi:10.1093/ije/dyh288 (2004).

42 Tenenbaum, D. & Maintainer, B. KEGGREST: Client-side REST access to the Kyoto Encyclopedia of Genes and Genomes (KEGG). R package version 1.38.0. (2022).

43 Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* **28**, 27-30, doi:DOI 10.1093/nar/28.1.27 (2000).

44 Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812-3814 (2003).

45 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).

46 van de Haar, J. *et al.* Identifying Epistasis in Cancer Genomes: A Delicate Affair. *Cell* **177**, 1375-1383, doi:10.1016/j.cell.2019.05.005 (2019).

47 Goldman, M. J. *et al.* Visualizing and interpreting cancer genomics data via the Xena platform. *Nature Biotechnology* **38**, 675-678, doi:10.1038/s41587-020-0546-8 (2020).

48 Cortes-Ciriano, I., Lee, S., Park, W. Y., Kim, T. M. & Park, P. J. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun* **8**, 15180, doi:10.1038/ncomms15180 (2017).

Acknowledgements:
The authors thank Minping Qian, Hong Wu, Zexian Zeng, Ruibin Xi, Ge Gao, Qing Li, Chao Tang, Jinzhi Lei, Yihan Lin, and Zemin Zhang for insightful discussions.

Funding: This work was supported by the National Key Research and Development Program of China (no. 2021YFF1200902) and the National Natural Science Foundation of China (nos. 32270689 and 12126305).

Author contributions: F.Z. performed data processing, data analysis, interpreted the data, and prepared the manuscript. M.D. acquired funding, supervised the study, and reviewed the manuscript.

Competing interests: The authors declare no competing interests.

Data and materials availability: The results published here are based on data generated by the TCGA Research Network (https://www.cancer.gov/tcga), the Genotype-Tissue Expression (GTEx) Project (https://www.gtexportal.org), UCSC Xena (https://xenabrowser.net/datapages), and the IARC database (http://ci5.iarc.fr/CI5-X/Pages/download.aspx). The analysis code is archived on GitHub (https://github.com/WhitePredator/sex-bias-in-cancer-initiation).

Supplementary Materials

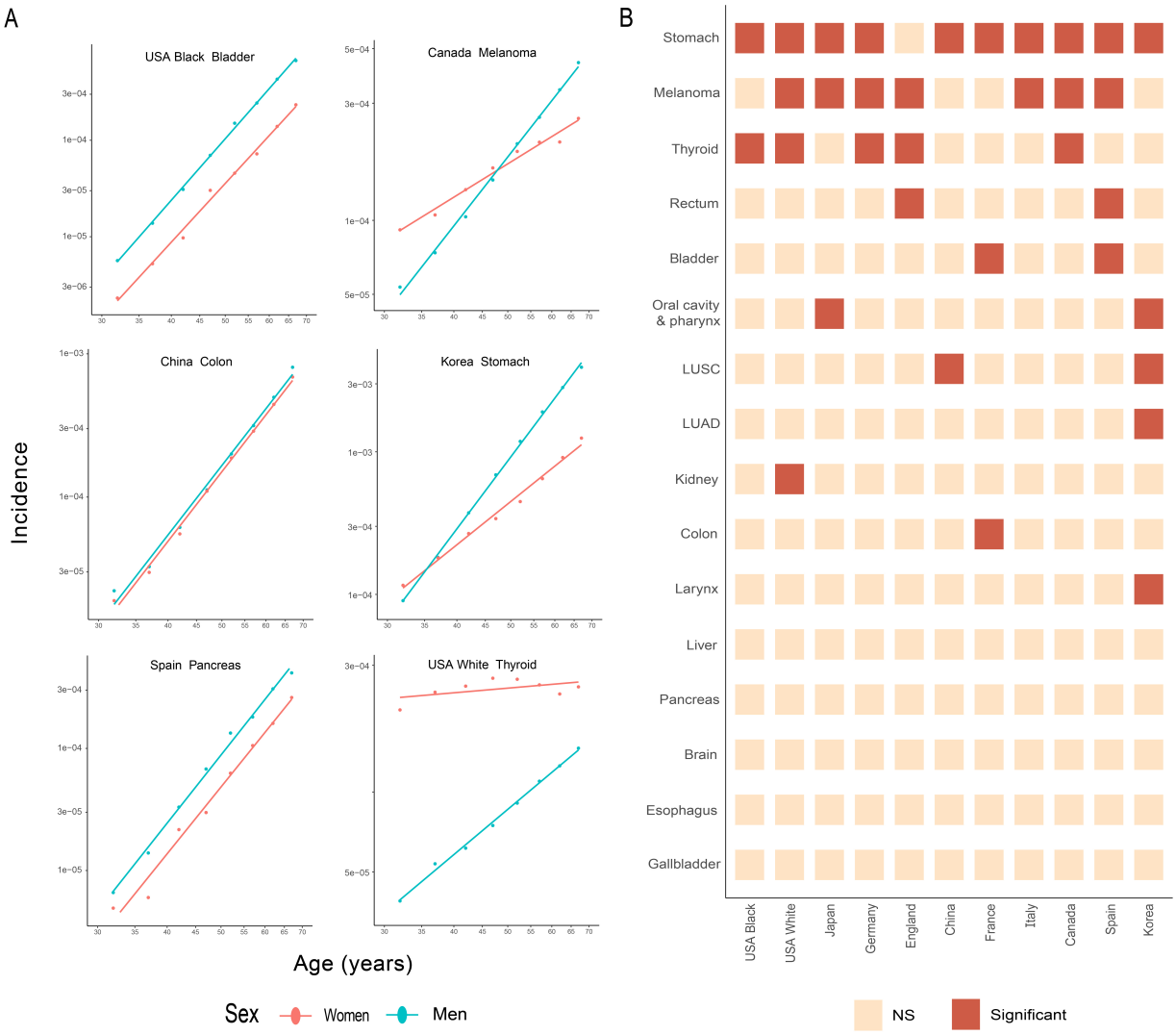
Materials and Methods
Figs. S1 to S7
Tables S1 to S7
References (39–48)
Data S1 to S5

Table 1. Sex-biased mutated genes.

Gene	Chromosome	Cancer	Female samples +/-	Male samples +/-	Relative risk (f/m)	LRT P value
CASP8	2	HNSC	23/108	19/332	3.24	3.85e-10
RBM10	X	LUAD	9/255	18/205	0.42	3.52e-4
SELP	1	SKCM	13/149	8/249	2.58	7.71e-3
SPAG17	1		17/145	13/244	2.07	7.71e-3
DNAH7	2	STAD	13/131	11/259	2.21	4.70e-2
LARP4B	10		13/131	8/262	3.05	4.27e-2
PGM5	9		13/131	8/262	3.05	4.27e-2
PTEN	10		14/130	12/258	2.19	4.70e-2

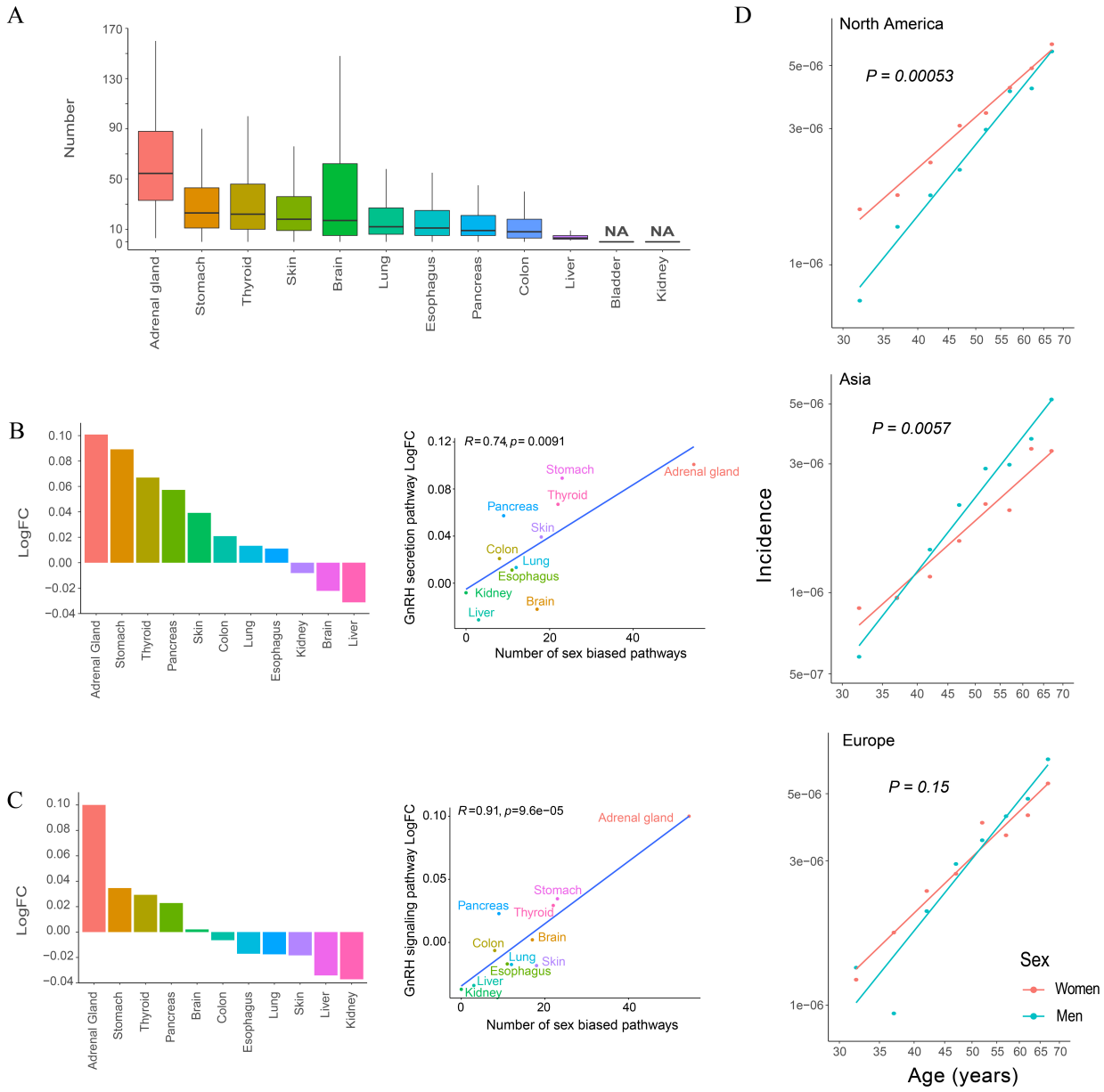
"Female samples +/-" and "Male samples +/-" refer to the numbers of samples from women or men that did or did not include a sex-biased gene mutation. "Relative risk (f/m)" represents the female-to-male relative risk of gene mutation frequency. "LRT P value" is the P value obtained from the likelihood ratio test, adjusted using the Benjamini-Hochberg metho

Fig. 1: Cancer incidence curves and prevalences of nonparallel patterns.



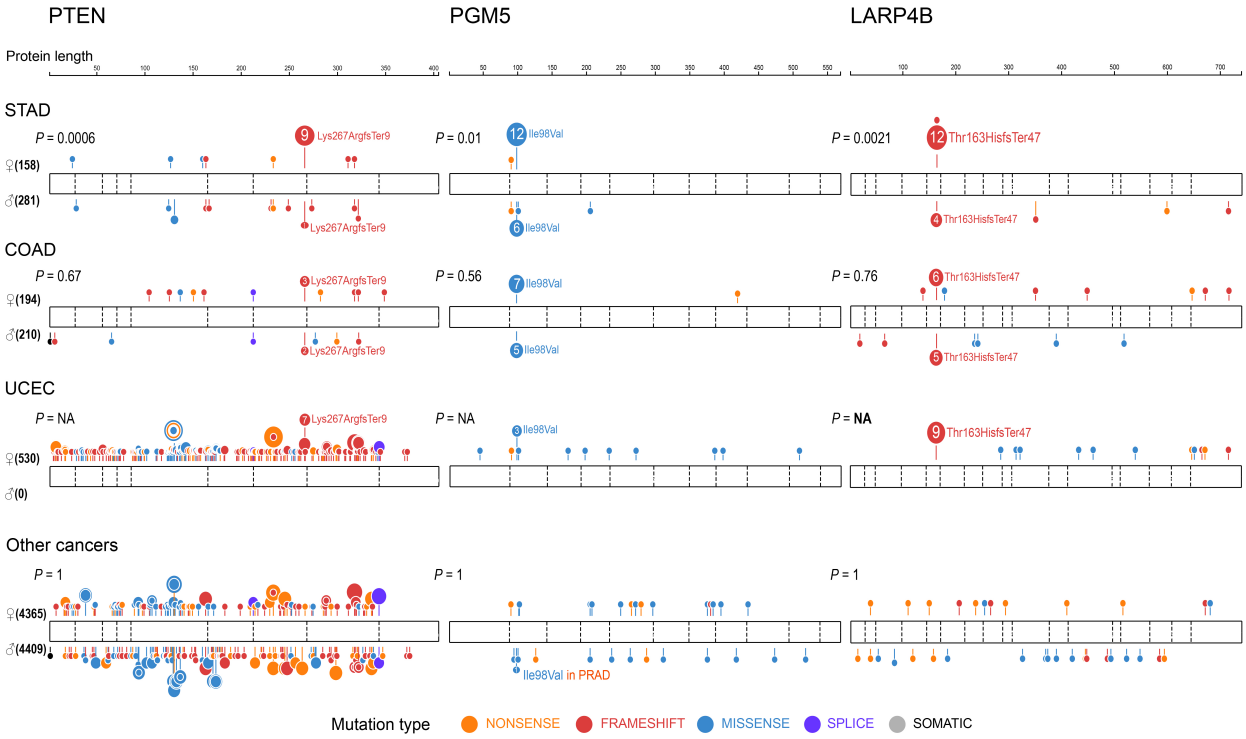
(A) Examples of parallel and nonparallel patterns. Both axes are on a logarithmic scale. Parallel patterns are shown on the left; nonparallel patterns are shown on the right. Incidences and fitted curves for women are shown in red; incidences and fitted curves for men are shown in blue. (B) Prevalences of significant nonparallel patterns in major cancer types according to region or ethnicity. Significant nonparallel patterns are shown in red; non-significant (NS) nonparallel patterns are shown in light yellow.

Fig. 2: Sex bias in tissue environments and prevalences of nonparallel patterns in adrenal gland cancer.



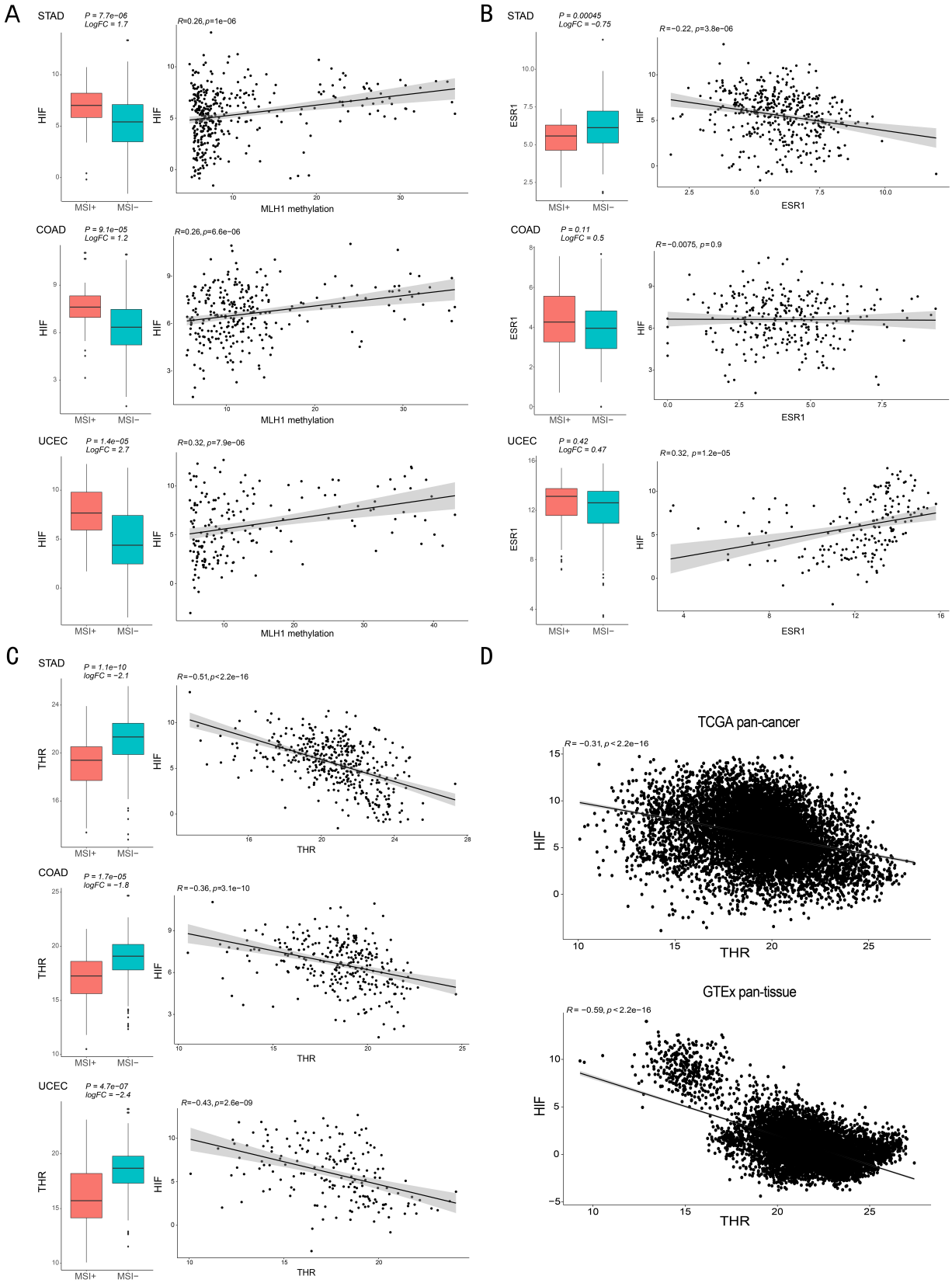
(A) Number of sex-biased pathways expressed in each tissue. The results of bladder and kidney analyses were regarded as not available (NA) because of small sample size. (B and C) Female-to-male log fold changes of GnRH secretion and GnRH signaling pathways among tissues. The results of bladder analyses were excluded because of small sample size. (D) Prevalences of nonparallel patterns in adrenal gland cancer. Regional cohorts within a single continent were merged to form a continental cohort.

Fig. 3: Mutational spectra of PTEN, PGM5, and LARP4B.



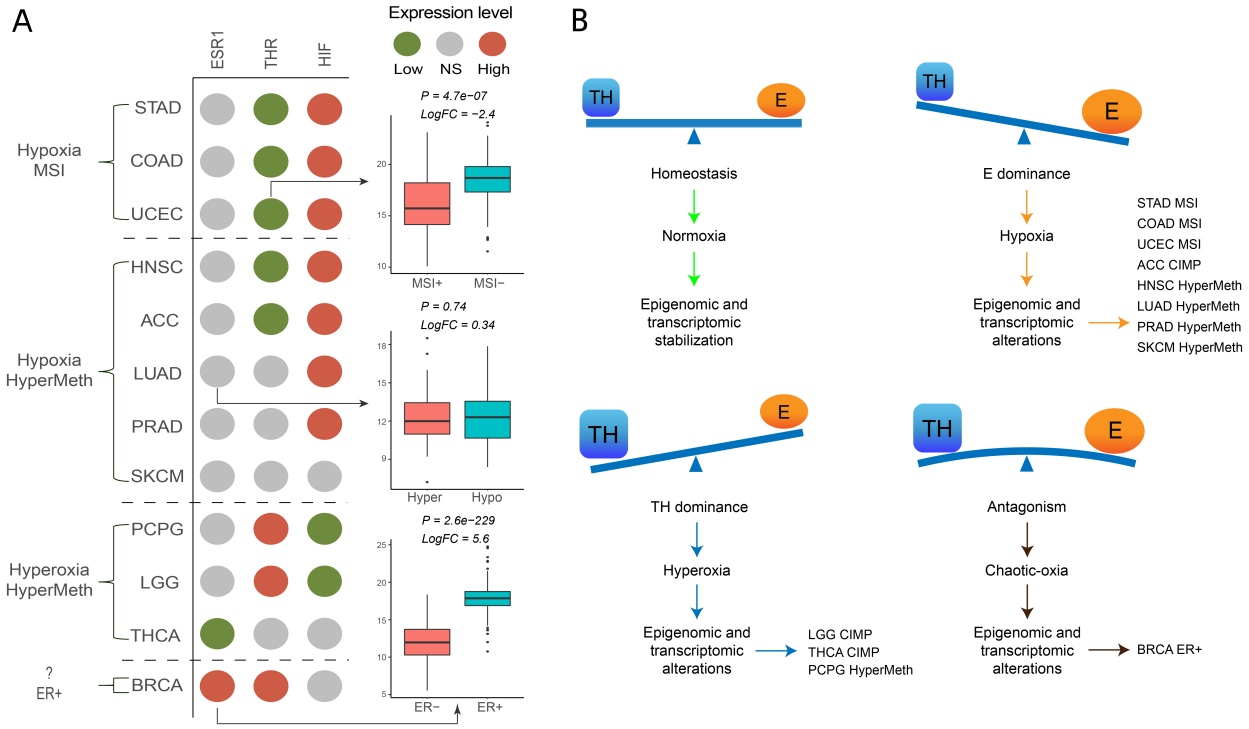
P values above gene body on the left denote statistical significance of hotspot sex bias. Because UCEC is only present in women, the P value is regarded as NA. The numbers of female and male samples for each cancer are indicated on the left in each row. Mutation types are shown in different colors and are annotated at the bottom of the figure.

Fig. 4: Relationships among HIF, ESR1, and THR.



(A) Comparison of HIF levels between MSI+ and MSI- samples, and correlations between HIF level and MLH1 methylation level across STAD, COAD, and UCEC. (B) Comparison of ESR1 levels between MSI+ and MSI- samples, and correlations between HIF level and ESR1 level across STAD, COAD, and UCEC. (C) Comparison of THR levels between MSI+ and MSI- samples, and correlations between HIF level and THR level across STAD, COAD, and UCEC. (D) Correlations between HIF level and THR level in pan-cancer and pan-tissue analyses. In all dot plots, *R* represents the Pearson correlation coefficient; P values were obtained using Student's t-test.

Fig. 5: HIF, ESR1, and THR expression signatures in TCGA cancers and the potential contributions of imbalances in estrogen and TH to carcinogenesis.



(A) Comparative expression signatures of ESR1, THR, and HIF are represented by colored dots. Red dots indicate significantly elevated expression, green dots indicate significantly reduced expression, and gray dots indicate no significant difference in expression when MSI+, HyperMeth, and ER+ cancers are compared with control cancers. Boxplots on the right are representative examples to illustrate the meaning of the dots. LGG: low-grade glioma, THCA: thyroid carcinoma. (B) The relationship between estrogen (E) and TH is illustrated by a seesaw diagram. In a state of homeostasis, the levels of E and TH are in balance. When this balance is disrupted, epigenomic and transcriptomic alterations can occur in affected tissues. These alterations can lead to changes in the tissue environment and function as a selective pressure, thereby inducing specific cancer subtypes.