

RESEARCH ARTICLE

Multi-context modeling of driver pathways reveals common and specific mechanisms across 23 cancer types

Wenjia Zhou^{1,2}, Junhua Zhang^{1*}

1 CEMS, NCMIS, RCSDS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, **2** School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, China

* zjh@amt.ac.cn

**OPEN ACCESS**

Citation: Zhou W, Zhang J (2025) Multi-context modeling of driver pathways reveals common and specific mechanisms across 23 cancer types. *PLoS Comput Biol* 21(8): e1013349. <https://doi.org/10.1371/journal.pcbi.1013349>

Editor: Melissa L. Kemp, Georgia Institute of Technology and Emory University, UNITED STATES OF AMERICA

Received: November 24, 2024

Accepted: July 21, 2025

Published: August 6, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1013349>

Copyright: © 2025 Zhou, Zhang. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Somatic mutation dataset in this study can be downloaded from the open platform including TCGA

Abstract

Discovery of cancer driver pathways is essential for targeted therapies, since these pathways govern tumor progression and treatment resistance. However, their context-specific patterns across populations remain poorly understood. Leveraging pan-cancer genomic data, we apply our two models, EntCDP and ModSDP, to perform stratified analyses from four perspectives: region, tumor type, age group, and risk factors. Our results reveal the regional biases in perturbed pathways, such as PI3K-Akt in Chinese patients and GPCR in American patients with bladder cancer. Subtype comparisons highlight the mTOR signaling in lung adenocarcinoma and the FoxO signaling in lung squamous cell carcinoma. Pediatric-adult comparisons emphasize the enrichment of Ras signaling in pediatric acute myeloid leukemia and PAK signaling in pediatric glioblastoma, respectively. Risk factor associations further link Notch-mediated pathways to alcohol consumption and CDKN-regulated pathways to obesity-related cancers. Our findings demonstrate the utility of stratified driver pathway analysis in uncovering common and specific mechanisms, which can help prioritize context-aware therapeutic targets.

Author summary

Cancer develops as a result of disruptions in signaling pathways driven by gene mutations. However, the heterogeneity of these pathways across patient groups with distinct clinical characteristics remains inadequately understood. Here, we aim to bridge the existing knowledge gap by utilizing our newly developed models: EntCDP and ModSDP. These models are applied to analyze gene mutation data and clinical information from 55 cohorts, encompassing 23 different cancer types across various platforms. We explore how geographic factors drive regional biases, uncover differences among cancers with similar functions or locations (e.g., digestive system cancers, lung subtypes, sex-specific cancers, and metastatic tumors with preferred sites), and examine age-related shifts

(<https://portal.gdc.cancer.gov/>), ICGC (<https://dcc.icgc.org/releases/current/Projects/>), PCAWG (<https://xenabrowser.net/datapages/?hub=https://pcawg.xenahubs.net:443>) and cBioPortal (<https://www.cbioportal.org/>). The code for the Matlab package of EntCDP and ModSDP is freely available at <https://github.com/zjh136/Project-EntCDP-ModSDP>.

Funding: This work was supported by the National Key Research and Development Program of China (2022YFA1004800 to JZ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

to reveal developmental and physiological impacts. Additionally, we investigate how lifestyle factors, such as smoking, alcohol consumption, and obesity, influence signaling pathways through environmental exposures. These findings offer a comprehensive framework for understanding the heterogeneity of cancer signaling across diverse populations and clinical contexts. Our study not only elucidates the mechanisms underlying cancer progression, but also identifies potential targets for group-based therapeutic strategies.

Introduction

Dysregulation of pathways associated with human cancers can lead to abnormal cell proliferation and tumorigenesis. One underlying mechanism driving this process is genetic mutations in key components of pathways [1,2]. High-throughput sequencing technologies have revealed abundant mutational landscapes across various types of cancer [3–5], with heterogeneity widely observed [6,7]. This heterogeneity is manifested by a larger number of genes that are infrequently altered in certain types of cancer, rather than by a small number of universal genes that are frequently altered across multiple tumors [8]. Therefore, a persistent challenge lies in identifying the driver genes that promote tumorigenesis from a multitude of mutant genes across different tissues, and further elucidating how these genes function at the level of signaling pathways.

Existing approaches for identifying driver pathways or modules can be classified based on interaction knowledge, mutation frequency and pathway characteristics, often utilizing optimization models or network-based methods [9,10]. In terms of the patterns through which genes constitute a signaling pathway, the concepts of coverage and mutual exclusivity were described as two key features. The underlying evidence is that tumors gain a sufficient selective advantage as long as a single alteration disrupts the same functional pathway. In this context, Vandin et al. [11] developed the Dendrix (de novo driver exclusivity) method to maximize a weight function characterized by coverage and exclusivity of pathways. Their Markov chain Monte Carlo (MCMC) approach was later improved by an efficient binary linear programming (BLP) model [12]. Furthermore, the ComMDP and SpeMDP [13] models were developed at the pan-cancer level to discover common gene sets shared across distinct tumor types and specific gene sets of certain one or multiple tumor types compared to other types, respectively. To address the limitations of ComMDP and SpeMDP in identifying cancer-associated signaling pathways, we recently developed two refined models: EntCDP (Entropy-based Common Driver Pathway) and ModSDP (Modified Specific Driver Pathway) [14]. EntCDP leverages information entropy to balance the trade-off between coverage and exclusivity in ComMDP. ModSDP improves upon SpeMDP by better accounting for exclusivity. These comparative analyses naturally pave the way for cancer research and personalized medicine: commonality can guide the adoption of a consistent drug regimen across different cancers, while specificity enables the development of targeted therapies for individual cancer types. Therefore, it is necessary to conduct pan-cancer research from different perspectives.

On the other hand, genes serve as carriers of signaling pathways and can lead to cancer through mutations that disrupt these pathways, but not all genes are associated with cancer. To identify oncogenic mutations capable of driving malignancy, Martínez et al. proposed the Integrative OncoGenomics (IntOGen) framework [15] and compiled a comprehensive compendium of 568 driver genes across most tumor types. However, there is a lack of understanding regarding how these genes function as components of signaling pathways. This inspired us

to narrow our search for signaling pathways to these driver genes, which are more biologically significant.

Many pan-cancer projects have provided extensive data on gene mutations and advanced the field of pan-cancer genome research [16–18]. However, few studies have utilized clinical information to compare the signaling pathways across multiple cancer types under different conditions. To address this gap, we collected gene mutation data as well as sample attributes of 23 types of cancer from multiple platforms (TCGA, ICGC, MSKCC, TARGET, etc.), and investigated common and specific pathways using EntCDP and ModSDP in patients classified based on the following four conditions: (1) For patients diagnosed with the same cancer but originating from different nations or regions (such as CN, AU, US, etc., 12 in total), we identified novel targeted drivers for treating the relevant cancers based on common findings, and highlighted therapeutic significance of pathway prediction derived from region-specific findings; (2) For cancers sharing similar functionality or anatomical location, we analyzed common and specific signaling characteristics among comparative cancers, including cancers of the digestive system, lung subtypes, sex-specific cancers, and cancers prone to metastasis along with their metastatic sites; (3) For children who developed malignant disease, we sought to identify biomarkers for subtypes of hematological and neural tumors. We also aimed to uncover the specific characteristics of glioblastoma multiforme and acute myeloid leukemia through comparisons with adults; (4) For patients with or without risk factors such as smoking, drinking and obesity, we aimed to determine the signaling pathways disrupted by these factors that trigger cancer. Furthermore, we sought to identify additional driver factors beyond the known risk factors. In summary, our study provides novel insights into targeted cancer therapies in different regions, ages, and risk factors, thereby advancing the field of personalized cancer treatment.

Results

Sample characteristics

To facilitate a comparative analysis of signaling pathways among multiple cancer types under different conditions, we collected tumor genomic mutation profiles from 55 cohorts representing 23 typical types of cancer. These cohorts comprised 14,564 adult and 1,539 pediatric tumor samples from large-scale cancer genomics programs, including The Cancer Genome Atlas (TCGA) [2], the International Cancer Genome Consortium (ICGC) [19] and cBioPortal [9] (Fig 1A and S1 Table). After the exclusion of silent mutations, we selected samples with at least one alteration event (e.g., somatic mutations, small insertions/deletions, copy number variations, etc.) for further analysis.

In addition, we collected clinical data for these samples to investigate clinicopathological characteristics, including region, age, smoking history, drinking history, and obesity. However, the paired clinicopathological data were not always complete and comprehensive (see **Materials and methods**). Among the 55 cohorts, 50 had nationality or regional attributes, while the remaining 5 cohorts were primarily sourced from the cBioPortal. For glioblastoma multiforme (GBM) and acute myeloid leukemia (AML) cohorts, all samples in cohorts LAML, GBM and GBM_MSKCC were from individuals aged 18 years or older, except for LAML_CN, which did not provide age information. To ensure that only adult samples were included in the LAML_KR cohort for comparison with the pediatric cohort, we excluded 3 patients under the age of 18 from the total of 205 samples. TCGA provided the most comprehensive exposure data for risk factor analysis: 96.48% (1426/1478) of the LUAD, HNSC and BLCA samples were used to examine the role of smoking as a risk factor; 96.46% (817/847) of the ESCA, HNSC and PAAD samples were used to assess the effects of drinking; and 66.18%

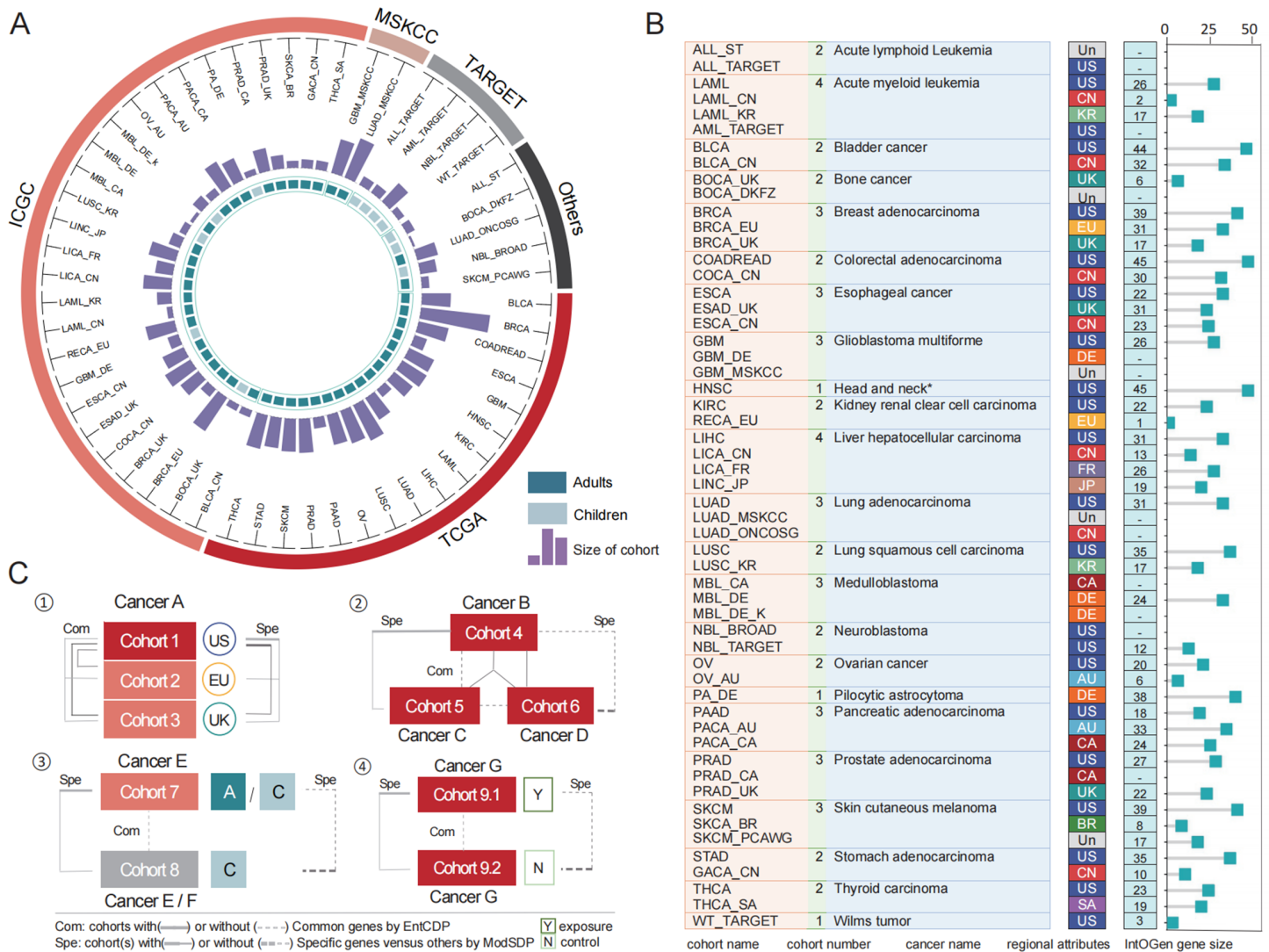


Fig 1. Datasets of genetic mutations and framework of analysis. (A) Summary of tumor genetic mutation datasets collected from public platforms. The outer ring indicates the sources of all cohorts included in this study, which are connected to their respective names through T-shaped crosses. The histogram inside illustrates the sample size of each cohort, and the innermost ring indicates whether the cohort mainly consists of adults or children. (B) Each column of the table provides detailed information about all 55 cohorts, including cohort names, the number of cohorts for each cancer type, cancer types, regional attributes, and the number of IntOGen genes for each cohort. The lollipop diagram on the right illustrates the numbers in the last column of the table. Different colors are used to distinguish different countries or areas. Un: unknown. -: no records. (C) Four parts of our analysis. Wires connecting different cohorts represent comparisons conducted using EntCDP (common pathways, Com) and ModSDP (specific pathways, Spe). The background colors of cohorts correspond to the platforms shown in (A). Head and neck*: Head and neck squamous cell carcinoma. A: adults; C: children; Y: exposure groups; N: control groups.

<https://doi.org/10.1371/journal.pcbi.1013349.g001>

(595/899) of the COADREAD and LIHC samples were included in the analysis of obesity-related risks.

We focused our search on a well-curated collection of 568 mutational cancer genes identified by IntOGen [15], which were derived from somatic mutations data of approximately 28,000 tumors across 66 cancer types. As nearly 80% (42 out of 55) of our cohorts can overlap those used in IntOGen (Fig 1B), the cancer genes identified by IntOGen within these cohorts serve as an important benchmark for validating our results. Therefore, we used EntCDP and ModSDP to identify and compare the driver pathways across 23 cancer types based on the

568 genes, providing insights into the associations between these pathways and region, age and risk factors (Fig 1C).

EntCDP and ModSDP reveal regional bias in driver genes of the same cancer

Robust identification of driver genes via pathway models. Before the investigation of pathway findings, we proposed a brief test to verify the reliability of the results derived from EntCDP and ModSDP models. The driver genes discovered by IntOGen provide insights into the mutational features of each cohort (Fig 1B and S1 Table) and provide a valuable benchmark for comparison with our findings. The main motivation is to examine, using the hypergeometric test, the degree of overlap between common (or specific) IntOGen genes of comparable cohorts and their common (or specific) driver gene sets discovered by EntCDP (or ModSDP) (Sect A in S1 Text). Since some cohorts lack IntOGen genes, we restricted the hypergeometric test to cohorts containing IntOGen genes (Sect B in S1 Text).

In spite of the incompleteness of IntOGen driver genes, those reported in the literature and associated with cohort characteristics, such as cancer type and geographic location, can serve as a credible reference for validating novel discoveries. In a sense, genes not assigned to any cohort by IntOGen and lacking annotation in the literature can be considered new candidate driver genes for a certain cancer.

Targeting novel driver genes in region-specific signaling pathways. The information on nationality or sequencing location in the collected data enables us to geographically stratify samples with the same cancer type (Fig 1B). Here, we focus on cohorts of each cancer type and apply EntCDP and ModSDP to cohorts of the same type but different geographic origins.

There is some overlap between the genes discovered by our models and those cohort-sensitive genes (i.e., IntOGen driver genes). At least 20% of common genes identified by EntCDP in each cohort group (Fig 2A) are IntOGen driver genes, all except kidney cohorts KIRC and RECA_EU, which share only one IntOGen gene (S2 Table). In addition, significant overlap between IntOGen and ModSDP can be seen in half (7/14) types of cancer (S3 Table).

While IntOGen overlapping driver genes can confirm the accuracy of our results, non-IntOGen driver genes can either be validated as drivers or predicted to serve in such a capacity. For example, two common genes, *TP53* and *PIK3CA*, identified in breast adenocarcinoma cohorts from the US, EU and UK, are IntOGen genes shared by all three cohorts (colored by dark purple in Fig 2A), while *GATA3* is a known IntOGen gene present in two of the cohorts (colored by light purple). The three genes function in the IGF1 receptor (IGF-1R) signaling pathway and are identified by EntCDP when the size of the search set (parameter *K*) is 3 (Fig 2C and Table S1.3 in S1 Data). It has been reported that the IGF-1R pathway plays an important role in gynecologic cancers such as breast tumors [20]. When *K* = 10, we identified other seven genes mutually exclusive with the three genes. Although *H3F3A*, *CARS*, *GTF2I* and *SEPT9* are not IntOGen genes of the three cohorts (colored by yellow), they have been proved to play important roles in the development of breast adenocarcinoma [21–24]. For example, *SEPT9*, a septin gene encoding multiple isoform, has been confirmed as a promising tool in breast cancer detection [24]. However, those not previously reported in the literature (colored by pink), including *TAF15*, *HIST1H4I* and *FAM46C*, may facilitate the identification of novel targets for the treatment of breast adenocarcinoma. Moreover, there is a series of driver genes specific to the cohort BRCA relative to BRCA_UK and BRCA_EU, with gene set {*CDH1*, *MAP2K4*, *PIK3R1*, *MAP3K1*} detected by ModSDP when

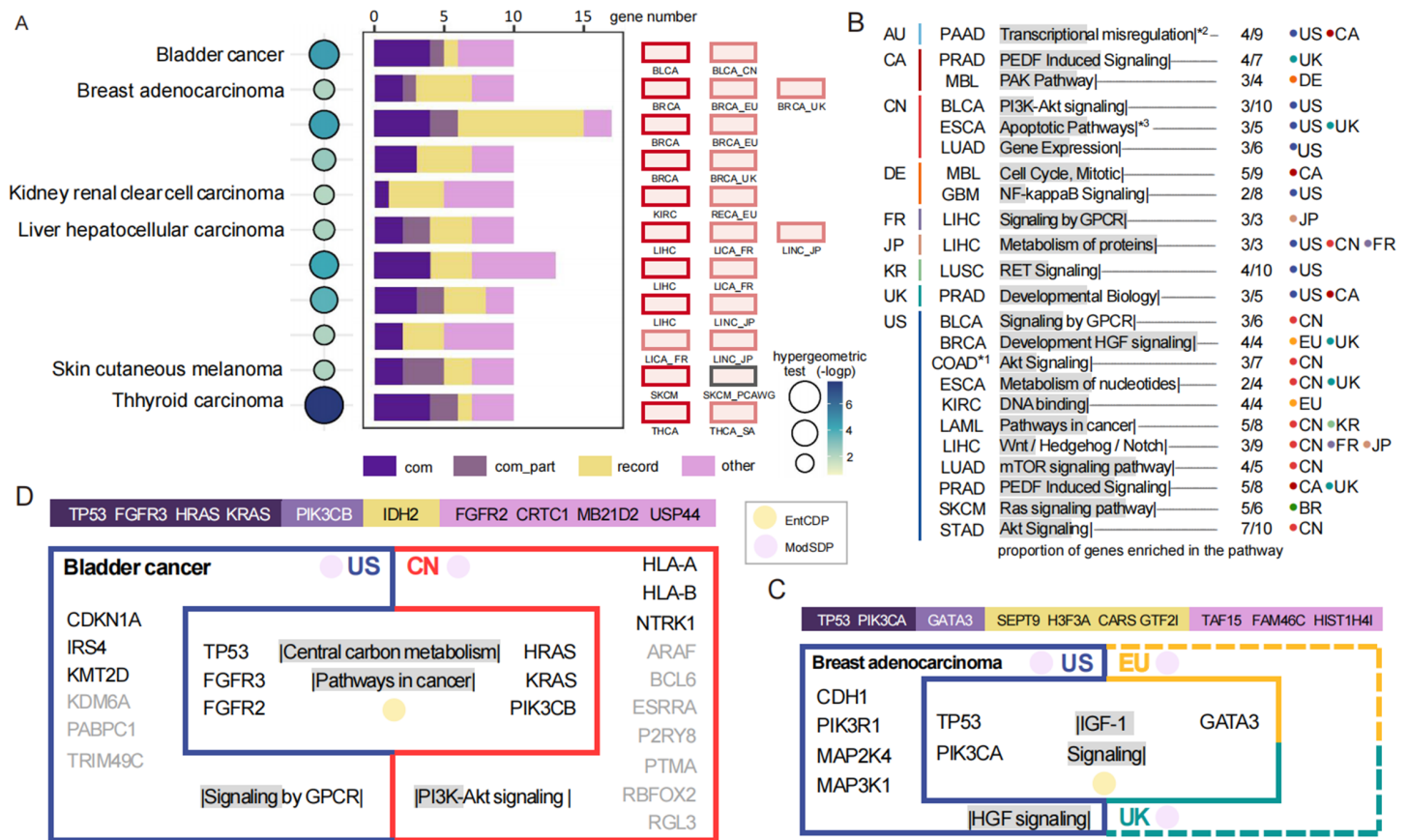


Fig 2. Results of the reliability test and regional differences in signaling pathways. (A) The hypergeometric test reveals significant results for six types of cancer with IntOGen genes in each cohort, which investigates the overlap between cohorts' common genes identified by IntOGen and EntCDP. The size of the circles corresponds to the significance level. Different colors in the bar graph present different features of genes identified by EntCDP: dark (light) purple indicates common genes obtained from all (part of) selected cohorts by IntOGen; yellow (pink) denotes non-IntOGen genes of any selected cohort that have (not) been verified as drivers in the literature. (B) Specific pathways of patients with a certain cancer type coming from one country (left) relative to other counties (right, denoted by colorful circles). The gray-filled bar for each signaling pathway indicates the proportion of identified genes enriched in that pathway, as shown by the ratio on the right. (C, D) Common signaling pathways among, and pathways specific to, patients with breast adenocarcinoma from the United States versus those from European Union countries and Britain (C), or with bladder cancer from the United States versus China (D). The dashed line indicates no significant results. The bar above clearly lists the names of genes with different characteristics from (A). *1: COADREAD, colorectal cancer; *2: Transcriptional misregulation in cancer; *3: Apoptotic pathways in synovial fibroblasts.

<https://doi.org/10.1371/journal.pcbi.1013349.g002>

$K = 4$ (Table S1.6 in S1 Data). The four genes are members of the development HGF signaling pathway (Fig 2B and 2C), which can provide valuable diagnostic, prognostic and predictive biomarkers for personalized cancer therapy [25]. Targeting this pathway may provide clinical benefits for breast cancer patients from the United States.

Significant overlap between common genes discovered by both IntOGen and EntCDP is also observed in bladder tumors ($p = 0.0001$). *FGFR3*, *HRAS* and *KRAS*, along with known oncogene *TP53*, are typical biomarkers for bladder cancer associated with relevant pathways, and are discovered by EntCDP when $K = 4$ (Table S1.1 in S1 Data). When $K = 10$, EntCDP detected another six genes mutually exclusive with these four genes. Among them, *PIK3CB* is an IntOGen gene of the bladder cohort from America but not China, and *IDH2* is not present in any cohort but has been reported to be of potential therapeutic significance [26]. The remaining four genes, *CRT1*, *MB21D2*, *USP44* and *FGFR2*, may drive bladder cancer through other mechanisms of pathway crosstalk, which requires further investigation.

A gene set comprising 10 genes is the only significant gene set specific to BLCA_CN compared to BLCA, among which *NTRK1*, *HLA-A* and *HLA-B* are associated with the PI3K-Akt signaling pathway (Fig 2C and 2D and Table S1.2 in S1 Data). In contrast, American patients with bladder cancer tend to harbor mutations that disrupt signaling by GPCR, with *CDKN1A*, *IRS4* and *KMT2D* identified in the 6-gene set. Apart from these genes in specific pathways, some light gray genes in Fig 2D, such as *ARAF*, *ESRRA* and *PTMA* (from the China cohort), as well as *KDM6A*, *PABPC1* and *TRIM49C* (from the American cohort), are predicted by ModSDP as region-specific drivers of bladder cancer and require further experimental validation.

Common pathways shared by patients from different countries can provide novel targeted drivers for cancer treatment. Meanwhile, the therapeutic significance of pathway prediction is embodied within specific findings distinguished among areas, as shown in Fig 2B. There are also differences in the specific signaling pathways among different cancer types in the same region, which emphasizes the need to distinguish both regional and cancer-type differences in cancer treatments.

EntCDP and ModSDP provide pathway analysis of tissues with similar locations or related functions

Generally, different types of tumors have distinct mutation profiles and therefore require different treatments. However, certain types of cancer share commonalities and connections, particularly in terms of function and anatomical location. To decipher potential commonalities and differences among various cancers, we primarily utilized mutation data from TCGA to mitigate the influence of confounding variables. We selected several comparable groups for analysis, including (1) esophageal carcinoma (ESCA) and gastric carcinoma (STAD); (2) lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD); (3) colorectal cancer (COADREAD) and liver hepatocellular carcinoma (LIHC); (4) breast adenocarcinoma (BRCA) and ovarian cancer (OV); and (5) kidney renal clear cell carcinoma (KIRC) and prostate adenocarcinoma (PRAD).

Signal in esophageal and gastric carcinoma. The esophagus and stomach, two crucial organs of the digestive system, are closely related. However, there is a lack of studies on tumor biomarkers through genotype similarity analysis. Therefore, understanding the etiologies of ESCA and STAD at the pathway level will facilitate screening and prevention efforts in high-risk populations.

By applying EntCDP to cohorts ESCA and STAD, we identified a common driver gene set when $K = 9$ (Fig 3A top and Table S2.6 in S1 Data). Among them, *ARID1A*, *TP53*, and *CDH1* are also IntOGen genes in both cohorts. Although not listed by IntOGen for these cohorts, *ACVR1* and *CXCR4* have been implicated in the tumorigenesis of both cancers [27,28]. The remaining genes, including *LDB1*, *PCBP1*, *CDK4*, and *UBE2D2*, are novel predictions by EntCDP and may inform future therapeutic strategies. Enrichment analysis further supports the relevance of these genes, since *TP53*, *CDH1*, *CDK4*, and *CXCR4* are typical representations of pathways in cancer (Fig 3B).

To investigate specific signals, we further applied ModSDP to identify distinct driver genes between cohorts ESCA and STAD. We identified *PRKARIA*, *SETD1B* and *TP53* as ESCA-specific driver genes, which are essential for the metabolism pathway. In contrast, *ARID1A*, *NFE2L2* and *TP53* were recognized as STAD-specific genes and serve as major drivers in the gastric cancer pathway (Fig 3A bottom and Table S2.7 in S1 Data). Notably, *TP53* plays a prominent role in the development of esophagogastric cancers, not only because of its high mutation frequency, but also due to its enrichment in multiple cancer-related

divergent prognostic outcomes [29]. Despite some success in the conventional treatment for lung cancer, subtype-specific targeted therapies remain necessary to understand genetic heterogeneity among subtypes.

The biological relevance of EntCDP is supported by the identification of the common gene set {*CTNNB1*, *KRAS*, *MET*, *PTPN6*, *TP53*} in cohorts LUAD and LUSC (Fig 3C top). Among them, *KRAS* is an IntOGen gene of cohort LUAD, whose mutation is also a critical criterion in targeted therapy for patients diagnosed with LUSC [30]. Additionally, this gene set is enriched in the proteoglycans in cancer pathway (Fig 3D). In a combined experiment [31], the expression of proteoglycan genes significantly decreased in tumors compared to normal lung tissue, aligning with our findings and suggesting their potential as promising biomarkers.

As for the specificity, we identified a LUAD-specific gene set: {*EGFR*, *STK11*, *PTPRD*}, among which *EGFR* and *STK11* are involved in the mTOR and PI3K-Akt signaling pathways (Fig 3C bottom). It has been reported that dioscin, a therapeutic agent for lung adenocarcinoma, can suppress the proliferation, invasion and EMT of lung cells via the inactivation of AKT/mTOR/GSK3 β signaling pathway [32]. Notably, *PTPRD* is a novel candidate predicted by ModSDP within LUAD-specific gene set and may represent a promising target for future experimental validation. By contrast, the LUSC-specific gene set includes {*KMT2D*, *NFE2L2*, *PTEN*, *SGK1*}. Among these, *KMT2D* and *NFE2L2* are key regulators of LUSC tumorigenesis [33,34] and well-established components of the FoxO pathway. Therefore, we predict that *PTEN* and *SGK1* may serve as effective therapeutic targets for LUSC by mediating the FoxO signaling pathway. Collectively, these findings demonstrate the ability of ModSDP to recover biologically meaningful and therapeutically relevant gene sets specific to cancer subtypes.

Typical malignant tumors in males and females. Sex differences in cancer arise from the effects of circulating sex hormones [35], which influence organs unique to or significantly different between males and females.

In the previous section, we focused on breast cancer due to its high prevalence among women in many countries. Here, we compared ovarian cancer and breast carcinoma given their close association [36]. We found no common gene sets between cohorts BRCA and OV; however, we surprisingly identified significant commonalities between cohorts BRCA and OV_AU ($p = 0.0016$). For $K = 8$, the ErbB signaling pathway emerged as a common pathway, characterized by *TP53*, *PIK3CA*, *RAC1* and *FGFR3* (Fig 3E top and 3F). Among these, *TP53* and *PIK3CA* are IntOGen genes of both cohorts, while the roles of the other six genes, *ACKR3*, *EIF1AX*, *FGFR3*, *LOX*, *RAC1*, and *GATA3*, have also been highlighted in studies of both ovarian cancer and breast carcinoma [37–47]. These findings highlight the biological relevance and robustness of the identified gene set.

For the differences, BRCA-specific analysis highlighted the T cell receptor (TCR) signaling pathway, whereas OV was specifically characterized by the TNFR1 pathway activation (Fig 3F). In BRCA, we identified *CARD11*, *PIK3CA*, *PTEN*, and *GATA3* as key genes in the TCR pathway when $K = 6$ (Fig 3E bottom). Gu-Trantien et al. examined the prognostic factor of CD4+Tfh cells and the function of the T cell receptor signaling pathway in human breast cancer [48]. Conversely, we detected {*TP53*, *CBL*, *NFKBIE*, *RBM10*} as OV-specific drivers, and the first three are components of the TNFR1 pathway that exerts indirect cytotoxic effects on ovarian cancer stem cells [49]. Furthermore, we identified novel predictions in these contexts, including *MED12* and *RABEP1* in BRCA, and *MED12* as an OV-specific candidate for further functional validation.

It seems that pathway analysis for the two male urological cancers is less specific, as the specific pathways identified are the pathways in cancer for KIRC and the transcriptional misregulation in cancer for PRAD, respectively. See Sect C in [S1 Text](#) for more details.

Beyond the four cancer type pairs detailed in our study, other comparable groups may also be informative in future clinical trials (Tables S2.11–S2.17 in [S1 Data](#)).

EntCDP and ModSDP display the signaling profile for children

In recent years, reports on pediatric tumors have increased rapidly and attracted widespread public attention. Although advanced technology has profoundly reduced the mortality rates of pediatric cancer, the threat of death and ineffective treatment continues to cast shadows over too many vulnerable children [50].

Focusing on prevalent childhood tumors, we analyzed both commonality and distinctions among different pediatric tumors and their distinctions from adult tumors. We first presented a comprehensive comparison between glioblastoma multiforme (GBM) and acute myeloid leukemia (AML), both of which pose threats to children and adults, from both cross-sectional and longitudinal perspectives (Fig 4A). Subsequently, we discussed the biomarkers for two subtypes of hematological tumors and four subtypes of neural tumors.

Comparison of adult and pediatric cohorts in AML and GBM. The prominence of GBM among brain tumors underscored by its high morbidity and mortality rates in both children and adults [50]. Meanwhile, AML, a representative liquid tumor, is the most frequently diagnosed leukemia in adults (25%) and accounts for 15–20% of cases in children [51].

By leveraging somatic mutation data from both adult and pediatric cohorts of GBM and AML, EntCDP successfully identifies biologically meaningful common gene sets between two age groups. Specifically, for GBM (adults) and GBM_DE (children), four genes (*CASP8*, *PIK3CA*, *PTEN*, *LPAR4*) related to pathways in cancer are identified (Fig 4A right A & C and Table S3.1 in [S1 Data](#)). For LAML (adults) and AML_TARGET (children), the common pathway is cytokine signaling in the immune system, comprising six genes: *TP53*, *FLT3*, *KIT*, *KRAS*, *NRAS*, and *PTPN11* (Fig 4A left A & C and Table S3.6 in [S1 Data](#)). This finding is consistent with the fact that abnormalities in cytokine signaling pathways are characteristic of all types of leukemia [52].

Age-specific differences in signaling pathways can guide the discovery of therapeutic targets tailored to pediatric and adult cancers. In GBM, *CEBPA* and *ID3* in the NF- κ B signaling pathway (Fig 4A right C/A) are discovered as specific genes of children compared to adults when $K = 8$ (Table S3.4 in [S1 Data](#)), which may serve as potential targets in childhood glioblastoma. The adult-specific signaling is the MAPK signaling pathway, regulated by *TP53*, *FGFR4*, *EGFR* and *IRAK1* (Fig 4A right A/C), which has been reported to promote glioblastoma malignancy [53]. Treatment for AML children may rely on downstream effectors of receptor tyrosine kinase (RTK), such as the Ras and phospholipase D signaling pathways [54], since we discovered *NRAS*, *KIT*, *PTPN11* and *KRAS* in the two pathways when comparing AML_TARGET (children) to LAML (adults), LAML_CN (adults) and LAML_KR (adults) at $K = 4$ (Fig 4A left C/A and Table S3.7 in [S1 Data](#)). When $K = 8$, we detected {*ARHGAP35*, *CEBPA*, *DNMT3A*, *ASXL1*, *PHF6*, *SUZ12*, *TNC*, *MTOR*} as an adult-specific gene set for cohort LAML (Table S3.10 in [S1 Data](#)), among which *CEBPA*, *MTOR*, *SUZ12* and *ARHGAP35* are core components of the developmental biology pathway (Fig 4A left A/C). All remaining genes except for *TNC* have been previously implicated in AML [55–57]. Thus, we predict *TNC* as a novel candidate gene identified by ModSDP.

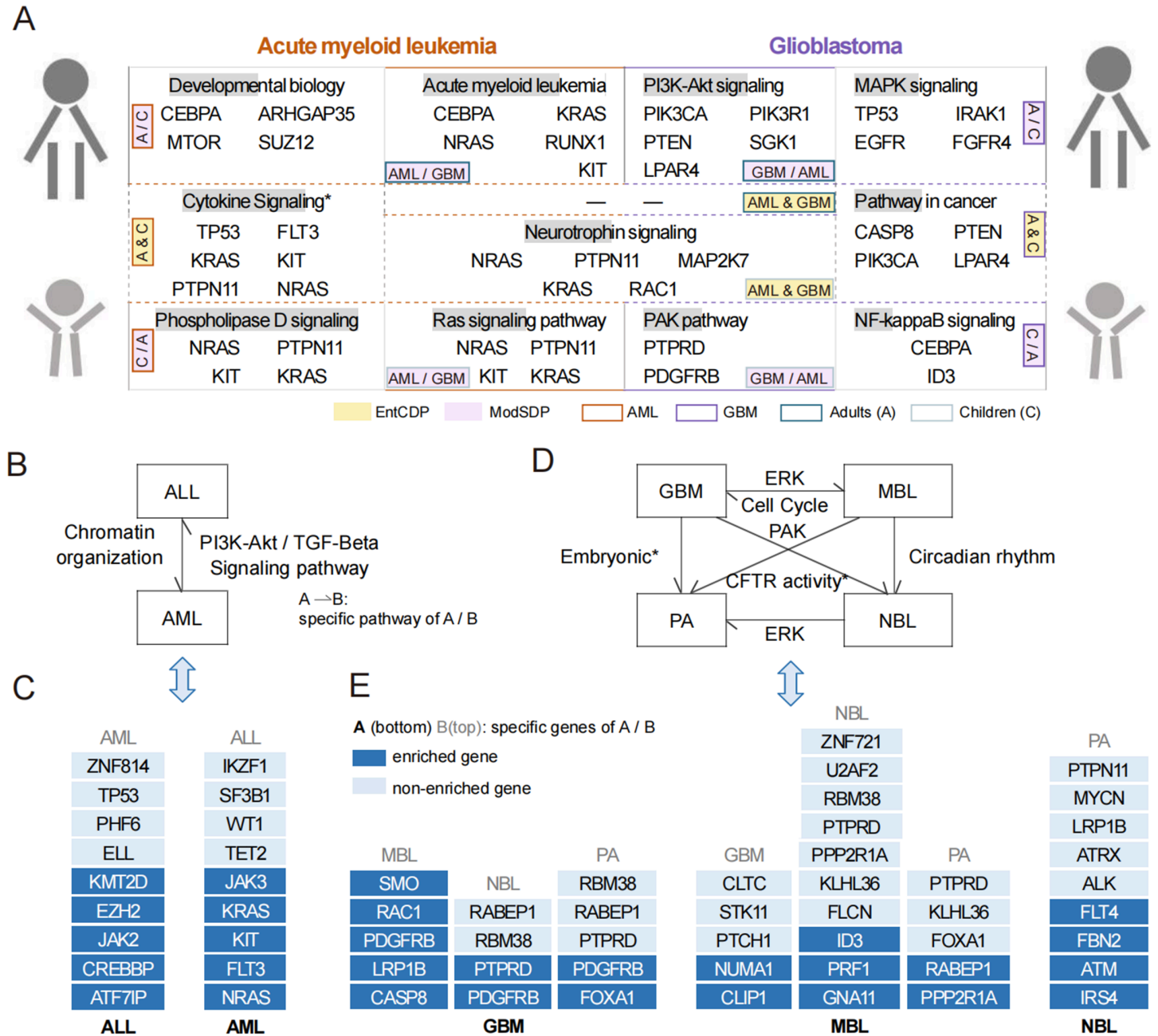


Fig 4. Interpretation of signaling pathways in pediatric and adult tumors. (A) Common and specific pathways of acute myeloid leukemia and glioblastoma in adults and children. (B, D) Specific signals of two hematopoietic tumors (B) and four neural tumors (D). The pathways labeled on directed edges represent significant signals specific to one cancer at the tail compared to another at the arrowhead. (C, E) Corresponding to (B) and (D), genes enriched in identified pathways are stacked from bottom to top and marked in dark blue. The remaining genes, depicted in light blue, are also identified by ModSDP but are not explicitly annotated for enrichment in the corresponding signaling pathway. Both dark blue and light blue genes are displayed under the same parameter *K*. Cytokine signaling*: Cytokine signaling in immune system; Embryonic*: Embryonic and induced pluripotent stem cells and lineage-specific markers; CFTR activity*: Regulation of CFTR activity.

<https://doi.org/10.1371/journal.pcbi.1013349.g004>

AML and GBM comparison within pediatric and adult groups. Next, we summarized some meaningful outcomes from the comparison between AML and GBM conducted separately in children and adults.

The results show that the neurotrophin signaling pathway [58] is the common factor driving tumor development in pediatric patients from the cohorts AML_TARGET and GBM_DE_K. However, no gene sets were detected from mutations in adults by EntCDP, which indicates greater cancer heterogeneity observed in adult patients (Fig 4A middle and Table S3.11 in S1 Data).

The distinct pathways identified between adult cancers demonstrate the biological significance of ModSDP. As illustrated in the upper part of Fig 4A, we successfully identified *CEBPA*, *NRAS*, *KIT*, *KRAS* and *RUNX1* as specific genes for adults with AML compared to those with GBM, and these genes are typical members of the acute myeloid leukemia pathway. On the other hand, adults with GBM are specifically affected by the PI3K-Akt signaling pathway (Table S3.12 in S1 Data), inhibition of which has been confirmed to be effective in clinical trials [59]. For children (lower part in Fig 4A), Ras and PAK signaling pathways are specifically associated with AML and GBM, respectively (Table S3.13 in S1 Data). As a tumor suppressor gene in the PAK pathway, inactivated *PTPRD* may lead to increased cell growth and suppression of apoptosis in GBM [60]. Mutations in the *PTPN11*, which encodes a phosphatase and regulates the Ras signaling pathway, are well characterized in children with AML [61,62]. Therefore, alterations in these two signaling pathways may have important implications for potential anticancer treatments.

In summary, the PI3K-Akt signaling pathway (*PIK3CA* and *PTEN*) is critical in adults with GBM, while the PAK signaling (*PTPRD* and *PDGFRB*) is specific to children with GBM. In AML, *KIT*, *NRAS*, and *KRAS* constitute core drivers, with *CEBPA* and *PTPN11* further distinguishing adult and pediatric cases, respectively. These genes represent key molecular drivers within their respective biological contexts. Other genes, such as *TNC* in adults with AML, and *CEBPA* and *ID3* in children with GBM, are proposed in our study as predicted therapeutic molecular targets within their associated signaling pathways.

Biomarkers for subtypes of pediatric hematological and neural tumors. The final section focuses on two subtypes of leukemia: acute lymphoid leukemia (ALL) and acute myeloid leukemia (AML), as well as four subtypes of neural tumors: glioblastoma (GBM), neuroblastoma (NBL), medulloblastoma (MBL) and pilocytic astrocytoma (PA).

Both ALL and AML are common malignancies in children and young adults, but they differ in biology characteristics and clinical responses to various chemotherapies [63]. When analyzing similarities between the two subtypes, we found a set of 10 common genes shared by ALL_ST and AML_TARGET (Table S3.19 in S1 Data), among which *TP53*, *NRAS*, *KIT*, *NF1*, *PTPN11* and *KRAS* are involved in the RET signaling. It is consistent with previous reports that AML cells exhibit activation of the RET signaling pathway [64]. Although there are few related reports for ALL, our results predict novel therapeutic targets, such as *WT1* and *FLT3*. Additionally, the two subtypes, ALL_TARGET and AML_TARGET, differ in signals dominated by specific genes shown in Fig 4B and 4C. Surprisingly, we noticed that *JAK2* and *JAK3*, two members of Janus kinase (JAK) family, are distinguishing markers for ALL and AML, respectively (Fig 4C).

Central nervous system tumors are the leading cause of cancer-related death in children [65]. In addition to GBM mentioned above, MBL and PA are particularly prevalent in the posterior fossa, and NBL is the most common extracranial solid tumor in children [66,67]. Despite much success, more targeted and effective therapies are needed for children whose tumors cannot be completely cured.

The commonality among different neural tumors can be observed only in the comparison between GBM and MBL (Table S3.23 in S1 Data). It is also worth noting that GBM and MBL cohorts have specific gene sets relative to other three brain tumors (Fig 4D and 4E). For

example, when compared to NBL cohorts (NBL_TARGET and NBL_BROAD), GBM_DE has specific genes, such as *PDGFRB* and *PTPRD* in the PAK pathway ($K = 4$), as well as *ID3* and *CEBPA* ($K = 8$) in the NF- κ B pathway (Table S3.25 in [S1 Data](#)), while MBL_DE_K and MBL_CA feature *ID3*, *GNA11*, and *PRF1* ($K = 10$) in circadian rhythm regulation (Table S3.28 in [S1 Data](#)). This suggested *ID3* could serve as a common regulatory gene for GBM and MBL. Unexpectedly, their common genes *PTPRD*, *PDGFRB*, *RBM38*, and *KLHL36* are also identified as overlapping specific genes distinguishing GBM and MBL from NBL. These results may provide insights into potential common targeted therapies.

Despite the similarities between GBM and MBL, we also focused on their molecular differences and successfully identified their specific therapeutic pathways using ModSDP. *CASP8*, *LRP1B*, *PDGFRB*, *RAC1* and *SMO* are representative specific genes for GBM_DE (Table S3.24 in [S1 Data](#)) and are involved in the ERK signaling. This pathway can inhibit the degradation of programmed death-ligand 1 (PD-L1) and maintain PD-L1 stability in glioblastoma [68]. On the other hand, we detected a specific gene set for MBL_CA: {*NUMA1*, *CLIP1*, *PTCH1*, *STK11*, *CLTC*}. Among these, *NUMA1* and *CLIP1* act in the cell cycle pathway, while *CLIP1*, *PTCH1*, *STK11*, and *CLTC* affect signaling by GPCR. This suggests that signaling pathway crosstalk may contribute to the development of MBL.

Subtype similarity is also embodied in NBL and PA, since the results of GBM versus PA are the same as those versus NBL when K ranges from 2 to 4 (Tables S3.25 and S3.26 in [S1 Data](#)). Compared with two NBL cohorts (NBL_TARGET and NBL_BROAD) and PA_DE, the specific genes of GBM_DE are {*CEBPA*, *SMO*} and {*FOXA1*, *RGL3*}, respectively. However, NBL exhibited relatively more heterogeneity than PA, because NBL (NBL_BROAD) has specific gene sets to PA (PA_DE), while PA lacks specific gene sets relative to any other subtype ([Fig 4D](#)). This specificity is observed in *IRS4*, *ATM*, *FBN2*, and *FLT4* ([Fig 4E](#) and Table S3.27 in [S1 Data](#), $K = 9$), which are involved in the ERK signaling pathway that has been verified to impact NBL progression [69]. The two MBL cohorts (MBL_DE_K and MBL_CA) also have specific genes compared to PA_DE (Table S3.29 in [S1 Data](#)). When $K = 5$, we detected {*FOXA1*, *KLHL36*, *PPP2R1A*, *PTPRD*, *RABEP1*}, among which *PPP2R1A* and *RABEP1* are involved in the regulation of CFTR activity pathway.

Overall, by comparing to PA, we identified subtype-specific pathways in neural tumors: PAK signaling (*PTPRD* and *PDGFRB*) in GBM, regulation of CFTR activity pathway (*PPP2R1A*, *RABEP1*) in MBL, and ERK signaling (*IRS4* and *ATM*) in NBL. In addition, based on the specificity relative to NBL, we predicted *RBM38*, *KLHL36*, and *ID3* as potential subtype-specific genes for GBM and MBL, while *LRP1B* and *PTPN11* appear to be specific to NBL. These predicted genes are promising for follow-up studies on their roles in signaling pathways.

Treatment for pediatric tumors should differ from that for adults, especially in light of mutations specific to children. The specificity among different tumor types also suggests the extent of tumor heterogeneity and the importance of prioritizing for targeted therapy. This underscores the role of personalized therapy, which can be achieved by careful stratification of patient populations.

EntCDP and ModSDP indicate that environmental factors may promote pathogenicity of cancers through pathways

Smoking, alcohol abuse and high BMI are three tumor-promoting agents that may induce genetic mutations and promote the development of cancers [70,71]. By comparing signaling pathways between smokers and nonsmokers, drinkers and nondrinkers as well as obese and normal-weight individuals with the same type of cancer, we analyzed the relationship

between the three factors and signaling pathways to inform targeted treatments for each group.

Risk factors differ between smokers and nonsmokers. It is well established that lung cancer is closely associated with tobacco use [70,72]. The fact that smokers outnumber nonsmokers in cohort LUAD further demonstrates the higher risk of lung cancer among smokers (Fig 5A).

To begin with, we analyzed the common driver pathways shared by all LUAD patients, regardless of their smoking history. Using EntCDP with $K = 9$, we detected {*KRAS*, *RAF1*, *TP53*, *MET*, *CTNNB1*, *PTPN6*, *FOXA1*, *MAX*, *NT5C3A*}. The first three genes are enriched in pathways related to non-small cell lung cancer, while the first four genes are enriched in the central carbon metabolism in cancer, a frequently altered pathway in lung adenocarcinoma [73]. These results demonstrate the effectiveness of EntCDP in capturing functionally meaningful genetic patterns.

However, what causes lung cancer in those patients who have never smoked remains unclear. Notably, when focusing on non-smoking LUAD patients, we identified signaling by GPCR when $K = 9$ (Fig 5 and Table S4.1 in S1 Data). There is reason to believe that *EGFR* plays a prominent role in inducing LUAD through this pathway, since it exhibits a high frequency of mutations and has been identified as a prognostic factor for lung cancer in never smokers (LCINS) [74]. In addition, we presume that the mutated *CYSLTR2* may be involved in lung cancer development through the production of aberrant GPCRs, based on the evidence observed in other cancers [75].

Human life is seriously affected by another smoking-associated cancer: head and neck squamous cell carcinoma (HNSC) [76], the sixth most common cancer worldwide [77]. There are significant differences in signaling pathways between HNSC patients exposed to tobacco and those who are not. In patients without a history of smoking, *RASA1*, *BRAF* and *CACNA1D* may promote the proliferation of head and neck cells by regulating the MAPK signaling pathway (S3A Fig and Table S4.2 in S1 Data). For smokers, we identified *COL1A1*, *PIK3CA*, and *LPAR4* as specific genes in the PI3K-Akt signaling pathway, which has been confirmed to be associated with HNSC [78,79]. We predicted the T-cell leukemia virus-1 (HTLV-1) infection pathway as a novel finding in HNSC because we identified *XPO1* and *EP300* as smoker-specific biomarkers involved in this pathway.

Shedding light on alcohol exposure. Beyond tobacco use, alcohol consumption is another major risk factor for head and neck cancer [80]. We found that drinkers generally carry specific genes such as *NOTCH1*, *CDKN2A* and *BRCA1* (S3C Fig and Table S4.4 in S1 Data). Their important role in the GPCR pathway is consistent with evidence that changes in GPCR function are implicated in ethanol exposure [81]. For nondrinkers, we detected {*CASP8*, *FH*, *MAPK1*, *RUNX1*, *STAT3*, *ETV5*, *EXT2*, *NRP1*, *ZNF626*} when $K = 9$. The first five genes are members of the pathways in cancer, among which *CASP8*, *STAT3* and *MAPK1* are related to viral carcinogenesis, including Kaposi's sarcoma-associated herpesvirus and human cytomegalovirus (HCMV) infection. This highlights the need to explore virus-related mechanisms in non-alcohol-associated tumor development.

In addition to HNSC, drinkers make up almost three-quarters of the patients in the cohort ESCA (Fig 5B). In particular, *NOTCH1*, characterized as a drinker-specific gene in HNSC, is also identified in ESCA drinkers, which underlines its core status in alcohol-related cancers [71]. *TP53* also attracts our attention because its high mutation rate exceeds 80% in both alcoholics and nonalcoholics. In addition, the remaining patients who never drink may be affected by a pathway called MHC class I antigen presentation (Fig 5B). The pathogenesis of

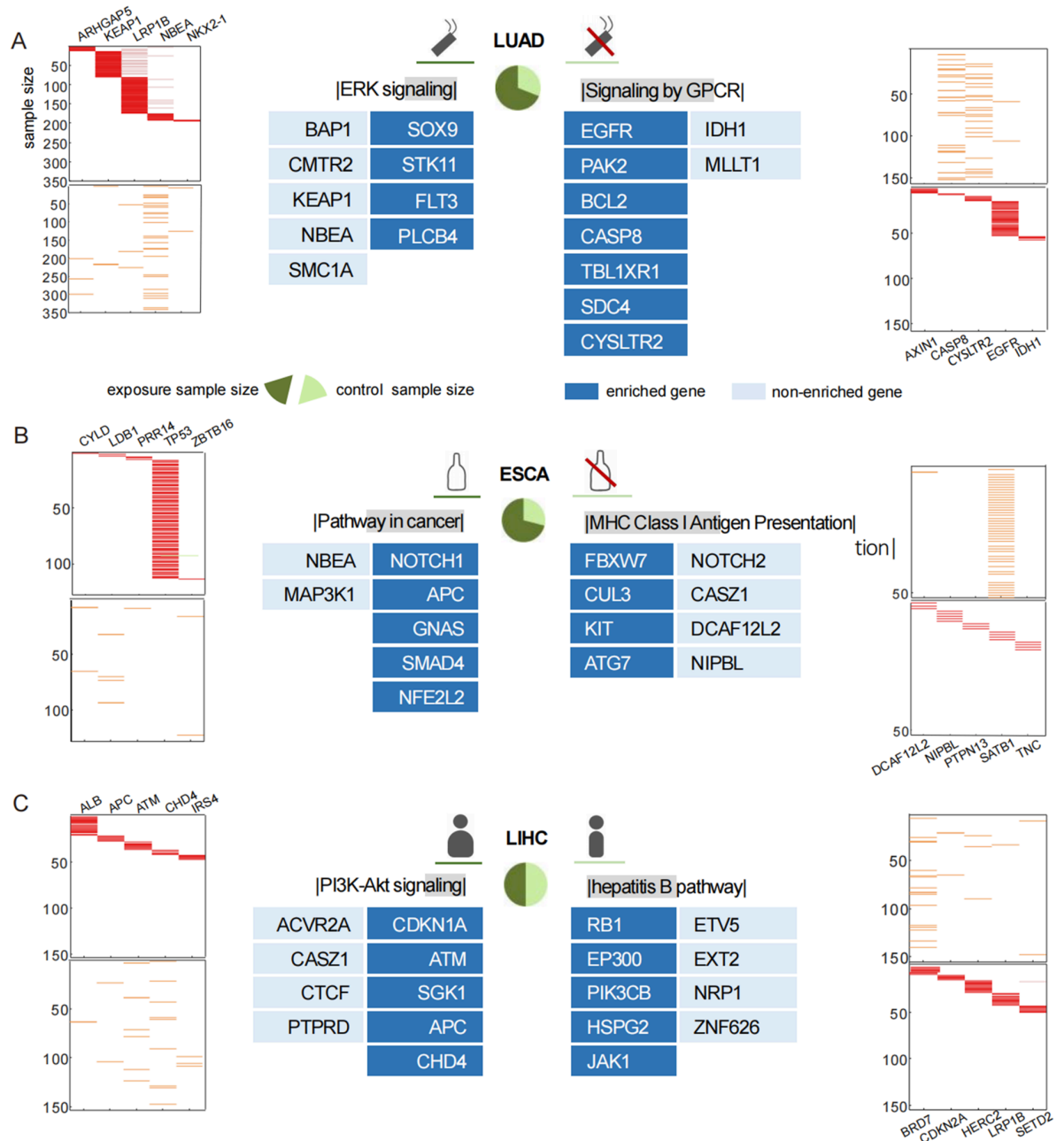


Fig 5. Three environmental agents promote tumor progression. Specific signaling pathways in patients with risk exposures to tobacco (A), alcohol (B) and high BMI (C) are illustrated on the left side, while the control groups without harmful habits or abnormal indices are shown on the right. The pie chart in the middle shows the proportions of the two groups. The heatmaps on both sides exhibit the gene mutation profiles of specific genes in the exposure group (the two maps on the left) and the control group (the two maps on the right) when parameter K is 5. Heatmaps in the same horizontal line share the same gene set. Genes enriched in identified pathways in the middle are filled in dark blue.

<https://doi.org/10.1371/journal.pcbi.1013349.g005>

eosinophilic esophagitis has been reported to depend on presenting antigens on the major histocompatibility complex (MHC) class II [82], which may imply the threat of MHC Class I antigen presentation in esophageal cancers.

Obesity as an underlying factor disrupting signaling pathways. It is believed that being overweight may also increase the risk of many cancers, such as stomach adenocarcinoma, liver hepatocellular carcinoma, breast carcinoma, and colorectal cancer [83]. Therefore, we investigated whether signaling pathways can serve as a bridge from obesity to cancer.

It has been reported that the PI3K/Akt/mTOR cascade is a target of many obesity-associated factors that regulate cell proliferation and survival [84]. We verified this result when we focused on patients with hepatocellular carcinoma. We identified *CDKN1A*, *ATM*, *SGK1*, *APC* and *CHD4* (Table S4.7 in [S1 Data](#), $K = 9$) as specific genes of overweight individuals. Enrichment analysis further revealed that these genes can be mapped to the PI3K-Akt signaling pathway. For those of normal weight, the hepatitis B pathway, composed of *RBI1*, *EP300*, *PIK3CB*, *HSPG2* and *JAK1* ($K = 9$), may serve as a key point for effective treatment and prevention ([Fig 5C](#)), because chronic infection with hepatitis viruses is a major causative factor for hepatocellular carcinoma [85].

Additionally, it is particularly common that people who are obese also tend to have colorectal cancer ([S3E Fig](#)). We conclude that mutation in *CDKN1B*, *CDKN2C*, *TP53*, *ATR*, and *CCR7* may increase the risk of developing colorectal cancer through the ERK signaling pathway in overweight individuals. However, those with a normal BMI should be cautious to prevent *Yersinia* infection [86], since *MAP2K1*, *PLCG1*, and *MYD88* exhibit mutual exclusivity and are significantly enriched in this pathway.

In summary, viral and bacterial infections as causes of human cancer have gained new significance because we identified the HCMV infection, HIV life cycle, hepatitis B, and *Yersinia* pathway in several control groups. Moreover, we have reason to believe that the Notch-mediated pathway may be responsible for bridging alcohol and cancers, while obesity would increase the risk of cancer through signals regulated by the CDKN gene family. For more details about risk factors for other cancers, please refer to [Sect D](#) in [S1 Text](#).

Discussion

All life on Earth, like the entire ecosystem in which we live, needs to maintain self-homeostasis through intricately complex molecular interactions. As is usually the case, forest fires are not attributed directly to global warming, even when a certain connection is apparent; instead, they are typically ascribed to specific triggers such as human activity and extreme weather events. Similar to the study of natural disasters, scientists are also thinking about the mechanisms of disease progression from local but critical events rather than the entire system. In this context, efforts are being made to compile a comprehensive list of disease-associated biomolecules, such as genes, proteins, mRNAs and related signaling pathways.

Currently, more attention is being paid to cancer, a severely complex disease and the leading cause of death worldwide. Since molecular targeted therapies represent a breakthrough in the management of this devastating illness, verification and functional annotation of driver genes are precious sources for treatment and are continually being supplemented and improved. We are involved in this major project and shift our perspective to the relationship between the signal molecular interrelation, namely, signaling pathway, and the causes of cancer. Therefore, we collect large-scale data on gene mutations in pan-cancer genomes from public databases and apply our two models, EntCDP and ModSDP, to identify common and specific pathways. Unlike traditional approaches that focus on a single cancer type, our comparative analysis enables the discovery of pathways shared by, or specific to, more than one type of cancer.

Our method addresses the shortcomings of the original models and improves the accuracy of signaling pathways identification. EntCDP solves the problem of inaccurate searches

caused by the large variations in mutation rates across different datasets, while ModSDP adjusts the coverage and exclusivity of signaling pathways to achieve a balanced consideration of both aspects [14]. At the same time, the signaling pathways we identify are constituted by more mutually exclusive alterations, which exert stronger selective pressure on tumors.

Among the tens of thousands of genes that a human cell carries, our identification scope is deliberately narrowed down to 568 reliable reference genes derived from an integrated model (IntOGen). Therefore, our results not only boost the efficiency of model execution, but also prevent genes not related to cancer from being falsely assumed to regulate oncogenic signals. Additionally, several cohort-sensitive genes identified by IntOGen are also recovered by our models, which supports the reliability of our findings.

In clinical practice, molecular subtypes can guide prognosis and therapeutic treatment in many cancer types. Beyond histological subtyping, we expect that individual patient information and clinical indicators can also be used as criteria for clinical decision-making, which is also the original intention of our study. Naturally, our two models have the advantage of categorical comparison.

On the one hand, common and specific signals among different types or subtypes of cancer can be easily identified for systematic analysis and interpretation. For example, EntCDP identified *TP53* as an important gene implicated in gastrointestinal disorders. When looking at women's cancers, disruption of the ErbB signaling pathway may contribute to breast and ovarian cancer. ModSDP identified the mTOR signaling pathway in lung squamous cell carcinoma and the FoxO signaling in lung adenocarcinoma, which can help with understanding subtype heterogeneity. On the other hand, we use the information provided by the data to subdivide samples of the same cancer according to the region, age, and risk factors. Region-specific analysis reveals that PI3K-Akt signaling pathway regulated by *HLA-A* and *HLA-B* is involved in Chinese patients with bladder cancer compared to Americans. In contrast, we predicted that therapeutic strategies for American patients might prioritize genes in signaling by GPCR, including *CDKN1A*, *IRS4*, and *KMT2D*. One ingenuity of our investigation in pediatric tumors is that we not only compare different cancers within the pediatric group, but also contrast children and adults with the same cancers (such as GBM and AML). Specifically, we found that the Ras signaling pathway, represented by *NRAS*, *KIT*, *PTPN11*, and *KRAS*, plays an important role in children with AML. Similarly, *CEBPA* and *ID3* are identified through NF- κ B signaling as potential drivers in pediatric GBM. Risk factor analysis underscores viral and bacterial infections (HIV, HBV, Yersinia) in cancer etiology. Additionally, we hypothesize that Notch signaling mediates alcohol-related carcinogenesis, while obesity may promote cancer via CDKN-regulated pathways.

We have considered further subdividing the data to provide a more precise reference for targeted therapy, but the application of our models is limited by the size of the data. In other words, our models can only search for more accurate signal systems with larger sample sizes. Specifically, limited sample sizes and lack of subtype data in pediatric patients are challenges to constructing significant and sophisticated mutation profiles for young patients, making it impossible to further account for the stage of tumor and subdivisions of childhood age. In addition, other divisions of datasets can also be explored. For example, it is also worth considering the races and continents not included in our study to better address the complexity and scope of the issues involved.

Our study highlights the need for classification and refinement of cancer patients. In addition to the reference lists provided in this paper, our models are applicable to other comparable populations. Finding the appropriate reference pathways and targets for clinical patients who best match their population marks a step forward in cancer genomics from bench to bedside.

Materials and methods

Data collection and processing

We obtained pan-cancer genomes across 23 tumor types from TCGA (<https://portal.gdc.cancer.gov/>), ICGC (<https://dcc.icgc.org/releases/current/Projects/>), cBioPortal (<https://www.cbioportal.org/>) and PCAWG (<https://xenabrowser.net/datapages/?hub=https://pcawg.xenahubs.net:443>). A total of 16,103 sequencing samples, consisting of 13,535 adult and 1,539 pediatric tumors, were collated into 55 datasets for analysis. We also collected clinical data for these samples to investigate clinicopathological characteristics, including region, age, smoking history, drinking history, and obesity.

Most TCGA cancer patients are adults, so we only selected patients aged 18 years or older to compare with children, while retaining their original cohort name. Similarly, for the pediatric cohorts, we eliminated samples older than 18, ensuring all pediatric patients included in our study are under 18 years old. According to the WHO criteria, individuals with a BMI greater than 25 were considered obese in our study.

EntCDP: Entropy-based common driver pathway identification model

EntCDP takes as input R gene mutation matrices $A_r = (a_{ij}^r)_{m_r \times G_r}$ ($r = 1, 2, \dots, R \geq 2$), where m_r is the number of samples and G_r is the number of genes in A_r . a_{ij}^r indicates the mutation status of the j -th gene in the i -th sample in the r -th matrix, with a value of 1 (mutation) or 0 (no mutation). First, the number of columns is expanded to $|G| = \bigcup_{r=1}^R G_r$, where G_r denotes the set of genes in A_r and $|\cdot|$ represents the cardinality, so that the same column in each matrix corresponds to the same gene. If a gene corresponding to a column is not present in a particular matrix, all entries in that column of the matrix are set to zero. To identify a common mutated driver gene set M containing K genes with large coverage and high exclusivity among the R cancers (i.e., M exhibits large coverage and high exclusivity in each cancer dataset A_r), EntCDP introduces an objective function $C_m(M)$:

$$C_m(M) = \omega \cdot \sum_{r=1}^R \left[2|\Gamma_{A_r}(M)| - \sum_{g \in M} |\Gamma_{A_r}(g)| \right], \quad (1)$$

where $\Gamma_{A_r}(g)$ denotes the set of samples in A_r with gene g mutated, and $|\Gamma_{A_r}(M)| = |\bigcup_{g \in M} \Gamma_{A_r}(g)|$ represents the coverage of M in A_r . In addition, $\omega = H(\{CR_r^{(M)}, r = 1, \dots, R\}) = -\sum_{r=1}^R (CR_r^{(M)} / \sum_{r=1}^R CR_r^{(M)}) \log(CR_r^{(M)} / \sum_{r=1}^R CR_r^{(M)})$, where the coverage rate $CR_r^{(M)} = |\Gamma_{A_r}(M)| / m_r$ and m_r is the sample size of A_r . As ω increases, the variation in the coverage of M across different cancer types decreases. We use genetic algorithms to maximize Eq 1. For more details about EntCDP, please refer to [14].

ModSDP: Modified specific driver pathway identification model

ModSDP takes as input two groups of gene mutation matrices: $A_r = (a_{ij}^r)_{m_r \times G_r}$ ($r = 1, 2, \dots, R$) and $B_t = (b_{ij}^t)_{m_t \times G_t}$ ($t = 1, 2, \dots, T$), where m_r and G_r are the number of samples and genes in A_r , and m_t and G_t are the number of samples and genes in B_t . Similarly, the number of columns of all matrices is expanded to the same, so that each column corresponds to the same gene across all matrices. To identify a specific mutated driver gene set M for the R cancer types relative to the T types (i.e., M exhibits large coverage and high exclusivity in each A_r but lacks these properties in each B_t), ModSDP introduces an objective function $S_m(M)$:

$$S_m(M) = \frac{1}{R} \sum_{r=1}^R \left[2|\Gamma_{A_r}(M)| - \sum_{g \in M} |\Gamma_{A_r}(g)| \right] - \frac{1}{T} \sum_{t=1}^T \left[K|\Gamma_{B_t}(M)| - \sum_{g \in M} |\Gamma_{B_t}(g)| \right] / (K-1), \quad (2)$$

where K is the number of genes in M . We maximize the binary form of Eq 2 using IBM ILOG CPLEX optimizer, a high-performance mathematical programming solver. Further details about ModSDP are provided in [14].

Permutation test

A gene set identified by EntCDP and ModSDP should be subjected to a permutation test to check whether it truly exhibits high coverage and high exclusivity characteristics of a signaling pathway. We compute two types of significance: individual significance, which evaluates a gene set within a single mutation matrix using statistics such as $2|\Gamma_{A_r}(M)| - \sum_{g \in M} |\Gamma_{A_r}(g)|$; and overall significance, which assesses a gene set across all matrices using the weights $C_m(M)$ in Eq 1 for EntCDP and $S_m(M)$ in Eq 2 for ModSDP.

The permutation test follows the approach in [12], where p -values are calculated based on 1,000 random permutations of the original mutation matrices. Each permutation preserves the mutation frequency of genes but randomized their sample assignments, thereby disrupting original mutation patterns. For EntCDP, significance is confirmed when both individual and overall p -values satisfy $p < 0.05$. For ModSDP, overall significance also requires $p < 0.05$, while individual significance requires $p < 0.05$ for each A_r and $p \geq 0.05$ for each B_t .

The results of each comparison are presented in individual tables in [S1 Data](#), listing significant driver gene sets containing 2 to 10 genes.

Pathway enrichment and gene set prioritization

We performed pathway enrichment analysis for each significant gene set by DAVID tool [87] to evaluate their potential enrichment in the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) databases. We set 0.05 as the significance threshold.

Subsequently, for each gene set enriched in significant pathways, we quantified the enrichment ratio as the maximum proportion of genes mapped to any enriched pathway. The gene set with the highest enrichment ratio was selected as the final candidate for downstream analysis.

Supporting information

S1 Text. Supporting information file including additional methods, analyses on regional bias, similar cancers, and risk factors.

(PDF)

S1 Fig. Venn diagrams highlight the overlap (dark region) in hypergeometric test for commonality between BRCA and BRCA_UK (A) and specificity of BRCA relative to BRCA_UK (B).

(TIF)

S2 Fig. Comparison of COADREAD and LIHC. (A) The result of common/specific gene sets of COADREAD and LIHC by EntCDP/ModSDP. (B) Regulation of genes involved in the signaling pathways shown in (A).

(TIF)

S3 Fig. Environmental agents promote tumor progression. Specific signaling pathways in patients with risk exposures to tobacco (A, B), alcohol (C, D) and high BMI (E) are illustrated on the left side, while control groups without harmful habits or abnormal indices are on the right.

(TIF)

S1 Table. Sample characteristics of cohorts.

(XLSX)

S2 Table. Driver genes identified by EntCDP and the comparison with those from IntOGen.

(XLSX)

S3 Table. Driver genes identified by ModSDP and the comparison with those from IntOGen.

(XLSX)

S1 Data. All tables of the results.

(PDF)

Author contributions

Conceptualization: Junhua Zhang.

Data curation: Wenjia Zhou.

Formal analysis: Wenjia Zhou, Junhua Zhang.

Funding acquisition: Junhua Zhang.

Investigation: Wenjia Zhou.

Methodology: Wenjia Zhou, Junhua Zhang.

Project administration: Junhua Zhang.

Resources: Wenjia Zhou, Junhua Zhang.

Software: Wenjia Zhou, Junhua Zhang.

Supervision: Junhua Zhang.

Visualization: Wenjia Zhou.

Writing – original draft: Wenjia Zhou, Junhua Zhang.

Writing – review & editing: Wenjia Zhou, Junhua Zhang.

References

1. Koni M, Pinnarò V, Brizzi MF. The Wnt signalling pathway: a tailored target in cancer. *Int J Mol Sci.* 2020;21(20):7697. <https://doi.org/10.3390/ijms21207697> PMID: 33080952
2. Hunter ZR, Yang G, Xu L, Liu X, Castillo JJ, Treon SP. Genomics, signaling, and treatment of Waldenström macroglobulinemia. *J Clin Oncol.* 2017;35(9):994–1001. <https://doi.org/10.1200/JCO.2016.71.0814> PMID: 28294689

3. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70. <https://doi.org/10.1038/nature11412> PMID: 23000897
4. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature*. 2020;578(7793):82–93. <https://doi.org/10.1038/s41586-020-1969-6> PMID: 32025007
5. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–4. <https://doi.org/10.1158/2159-8290.CD-12-0095> PMID: 22588877
6. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*. 2013;501(7467):338–45. <https://doi.org/10.1038/nature12625> PMID: 24048066
7. Reiter JG, Baretti M, Gerold JM, Makohon-Moore AP, Daud A, Iacobuzio-Donahue CA, et al. An analysis of genetic heterogeneity in untreated cancers. *Nat Rev Cancer*. 2019;19(11):639–50. <https://doi.org/10.1038/s41568-019-0185-x> PMID: 31455892
8. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science*. 2013;339(6127):1546–58. <https://doi.org/10.1126/science.1235122> PMID: 23539594
9. Leiserson MDM, Vandin F, Wu H-T, Dobson JR, Eldridge JV, Thomas JL, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet*. 2015;47(2):106–14. <https://doi.org/10.1038/ng.3168> PMID: 25501392
10. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res*. 2012;22(2):398–406. <https://doi.org/10.1101/gr.125567.111> PMID: 21908773
11. Vandin F, Upfal E, Raphael BJ. De novo discovery of mutated driver pathways in cancer. *Genome Res*. 2012;22(2):375–85. <https://doi.org/10.1101/gr.120477.111> PMID: 21653252
12. Zhao J, Zhang S, Wu L-Y, Zhang X-S. Efficient methods for identifying mutated driver pathways in cancer. *Bioinformatics*. 2012;28(22):2940–7. <https://doi.org/10.1093/bioinformatics/bts564> PMID: 22982574
13. Zhang J, Zhang S. Discovery of cancer common and specific driver gene sets. *Nucleic Acids Res*. 2017;45(10):e86. <https://doi.org/10.1093/nar/gkx089> PMID: 28168295
14. Zhou W, Zhang J. Pan-cancer study on transition of signaling systems from primary to metastatic tumors. *bioRxiv*. 2024. <https://doi.org/https://doi.org/10.1101/2024.111619269>
15. Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. *Nat Rev Cancer*. 2020;20(10):555–72. <https://doi.org/10.1038/s41568-020-0290-x> PMID: 32778778
16. Wang N, He D-N, Wu Z-Y, Zhu X, Wen X-L, Li X-H, et al. Oncogenic signaling pathway dysregulation landscape reveals the role of pathways at multiple omics levels in pan-cancer. *Front Genet*. 2022;13:916400. <https://doi.org/10.3389/fgene.2022.916400> PMID: 36061170
17. Wu J, Pan K, Zhu K, Cai Q. Identifying Common Driver Pathways based on Pan-cancer Data. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2019. p. 542–7. <https://doi.org/10.1109/bibm47256.2019.8983414>
18. Wu J, Pan K, Li G, Zhu K, Cai Q. Two novel models and a parthenogenetic algorithm for detecting common driver pathways from pan-cancer data. *Engineering Applications of Artificial Intelligence*. 2020;96:104010. <https://doi.org/10.1016/j.engappai.2020.104010>
19. Moloney JN, Cotter TG. ROS signalling in the biology of cancer. *Semin Cell Dev Biol*. 2018;80:50–64. <https://doi.org/10.1016/j.semdb.2017.05.023> PMID: 28587975
20. Bruchim I, Werner H. Targeting IGF-1 signaling pathways in gynecologic malignancies. *Expert Opin Ther Targets*. 2013;17(3):307–20. <https://doi.org/10.1517/14728222.2013.749863> PMID: 23294364
21. Fuentes-Antras J, Guevara-Hoyer K, Baliu-Piqué M, García-Sáenz JA, Pérez-Segura P, Pandiella A. Adoptive cell therapy in breast cancer: a current perspective of next-generation medicine. *Frontiers in Oncology*. 2020;10:605633.
22. Luangxay T, Osako T, Yonekura R, Sugiura Y, Kikuchi M, Gomi N, et al. Giant cell tumor of soft tissue of the breast: case report with H3F3A mutation analysis and review of the literature. *Pathol Res Pract*. 2020;216(2):152750. <https://doi.org/10.1016/j.prp.2019.152750> PMID: 31784095
23. Li K, Xu C, Du Y, Junaid M, Kaushik A-C, Wei D-Q. Comprehensive epigenetic analyses reveal master regulators driving lung metastasis of breast cancer. *J Cell Mol Med*. 2019;23(8):5415–31. <https://doi.org/10.1111/jcmm.14424> PMID: 31215771
24. Connolly D, Yang Z, Castaldi M, Simmons N, Oktay MH, Coniglio S, et al. Septin 9 isoform expression, localization and epigenetic changes during human and mouse breast cancer progression. *Breast Cancer Res*. 2011;13(4):R76. <https://doi.org/10.1186/bcr2924> PMID: 21831286

25. Moosavi F, Giovannetti E, Saso L, Firuzi O. HGF/MET pathway aberrations as diagnostic, prognostic, and predictive biomarkers in human cancers. *Crit Rev Clin Lab Sci*. 2019;56(8):533–66. <https://doi.org/10.1080/10408363.2019.1653821> PMID: 31512514
26. Roy S, Pradhan D, Ernst WL, Mercurio S, Najjar Y, Parikh R, et al. Next-generation sequencing-based molecular characterization of primary urinary bladder adenocarcinoma. *Mod Pathol*. 2017;30(8):1133–43. <https://doi.org/10.1038/modpathol.2017.33> PMID: 28548125
27. Sun Z, Liu C, Jiang WG, Ye L. Deregulated bone morphogenetic proteins and their receptors are associated with disease progression of gastric cancer. *Comput Struct Biotechnol J*. 2020;18:177–88. <https://doi.org/10.1016/j.csbj.2019.12.014> PMID: 31988704
28. Tan Y, Lu X, Cheng Z, Pan G, Liu S, Apizajai P, et al. miR-148a regulates the stem cell-like side populations distribution by affecting the expression of ACVR1 in esophageal squamous cell carcinoma. *Onco Targets Ther*. 2020;13:8079–94. <https://doi.org/10.2147/OTT.S248925> PMID: 32904700
29. Relli V, Trerotola M, Guerra E, Alberti S. Abandoning the notion of non-small cell lung cancer. *Trends Mol Med*. 2019;25(7):585–94. <https://doi.org/10.1016/j.molmed.2019.04.012> PMID: 31155338
30. Ito M, Codony-Servat C, Codony-Servat J, Lligé D, Chaib I, Sun X, et al. Targeting PKCL-PAK1 signaling pathways in EGFR and KRAS mutant adenocarcinoma and lung squamous cell carcinoma. *Cell Commun Signal*. 2019;17(1):137. <https://doi.org/10.1186/s12964-019-0446-z> PMID: 31660987
31. Rangel MP, de Sá VK, Prieto T, Martins JRM, Olivieri ER, Carraro D, et al. Biomolecular analysis of matrix proteoglycans as biomarkers in non small cell lung cancer. *Glycoconj J*. 2018;35(2):233–42. <https://doi.org/10.1007/s10719-018-9815-x> PMID: 29502190
32. Mao W, Yin H, Chen W, Zhao T, Wu S, Jin H, et al. Network pharmacology and experimental evidence reveal dioscin suppresses proliferation, invasion, and EMT via AKT/GSK3b/mTOR signaling in lung adenocarcinoma. *Drug Des Devel Ther*. 2020;14:2135–47. <https://doi.org/10.2147/DDDT.S249651> PMID: 32546976
33. Pan Y, Han H, Hu H, Wang H, Song Y, Hao Y, et al. KMT2D deficiency drives lung squamous cell carcinoma and hypersensitivity to RTK-RAS inhibition. *Cancer Cell*. 2023;41(1):88–105.e8. <https://doi.org/10.1016/j.ccell.2022.11.015> PMID: 36525973
34. Paik PK, Fan PD, Qeriqi B, Namakydoust A, Daly B, Ahn L, et al. Targeting NFE2L2/KEAP1 mutations in advanced NSCLC with the TORC1/2 inhibitor TAK-228. *Journal of Thoracic Oncology*. 2023;18(4):516–26.
35. Rubin JB, Lagas JS, Broestl L, Sponagel J, Rockwell N, Rhee G, et al. Sex differences in cancer mechanisms. *Biol Sex Differ*. 2020;11(1):17. <https://doi.org/10.1186/s13293-020-00291-x> PMID: 32295632
36. Zografos GC, Panou M, Panou N. Common risk factors of breast and ovarian cancer: recent view. *Int J Gynecol Cancer*. 2004;14(5):721–40. <https://doi.org/10.1111/j.1048-891X.2004.14503.x> PMID: 15361179
37. Neves M, Fumagalli A, van den Bor J, Marin P, Smit MJ, Mayor F. The role of ACKR3 in breast, lung, and brain cancer. *Mol Pharmacol*. 2019;96(6):819–25. <https://doi.org/10.1124/mol.118.115279> PMID: 30745320
38. Ha JH, Jayaraman M, Yan M, Dhanasekaran P, Isidoro C, Song YS, et al. GNAi2/gip2-regulated transcriptome and its therapeutic significance in ovarian cancer. *Biomolecules*. 2021;11(8):1211. <https://doi.org/10.3390/biom11081211> PMID: 34439877
39. Etemadmoghadam D, Azar WJ, Lei Y, Moujaber T, Garsed DW, Kennedy CJ, et al. EIF1AX and NRAS mutations co-occur and cooperate in low-grade serous ovarian carcinomas. *Cancer Res*. 2017;77(16):4268–78. <https://doi.org/10.1158/0008-5472.CAN-16-2224> PMID: 28646021
40. Li Y, Guo L, Ying S, Feng G-H, Zhang Y. Transcriptional repression of p21 by EIF1AX promotes the proliferation of breast cancer cells. *Cell Prolif*. 2020;53(10):e12903. <https://doi.org/10.1111/cpr.12903> PMID: 32926483
41. Chew NJ, Nguyen EV, Su S-P, Novy K, Chan HC, Nguyen LK, et al. FGFR3 signaling and function in triple negative breast cancer. *Cell Commun Signal*. 2020;18(1):13. <https://doi.org/10.1186/s12964-019-0486-4> PMID: 31987043
42. Saatci O, Kaymak A, Raza U, Ersan PG, Akbulut O, Banister CE, et al. Targeting lysyl oxidase (LOX) overcomes chemotherapy resistance in triple negative breast cancer. *Nat Commun*. 2020;11(1):2416. <https://doi.org/10.1038/s41467-020-16199-4> PMID: 32415208
43. Ye M, Zhou J, Gao Y, Pan S, Zhu X. The prognostic value of the lysyl oxidase family in ovarian cancer. *J Clin Lab Anal*. 2020;34(12):e23538. <https://doi.org/10.1002/jcla.23538> PMID: 33058284

44. Takaku M, Grimm SA, Wade PA. GATA3 in breast cancer: tumor suppressor or oncogene?. *Gene Expr.* 2015;16(4):163–8. <https://doi.org/10.3727/105221615X14399878166113> PMID: 26637396
45. El-Arabey AA, Denizli M, Kanlikilicer P, Bayraktar R, Ivan C, Rashed M, et al. GATA3 as a master regulator for interactions of tumor-associated macrophages with high-grade serous ovarian carcinoma. *Cell Signal.* 2020;68:109539. <https://doi.org/10.1016/j.cellsig.2020.109539> PMID: 31935430
46. Yamaguchi M, Takagi K, Sato A, Miki Y, Miyashita M, Sasano H, et al. Rac1 activation in human breast carcinoma as a prognostic factor associated with therapeutic resistance. *Breast Cancer.* 2020;27(5):919–28. <https://doi.org/10.1007/s12282-020-01091-2> PMID: 32314182
47. Hudson LG, Gillette JM, Kang H, Rivera MR, Wandinger-Ness A. Ovarian tumor microenvironment signaling: convergence on the Rac1 GTPase. *Cancers (Basel).* 2018;10(10):358. <https://doi.org/10.3390/cancers10100358> PMID: 30261690
48. Gu-Trantien C, Loi S, Garaud S, Equeter C, Libin M, de Wind A, et al. CD4+ follicular helper T cell infiltration predicts breast cancer survival. *J Clin Invest.* 2013;123(7):2873–92. <https://doi.org/10.1172/JCI67428> PMID: 23778140
49. Jo E, Jang H-J, Yang KE, Jang MS, Huh YH, Yoo H-S, et al. Cordyceps militaris induces apoptosis in ovarian cancer cells through TNF- α /TNFR1-mediated inhibition of NF- κ B phosphorylation. *BMC Complement Med Ther.* 2020;20(1):1. <https://doi.org/10.1186/s12906-019-2780-5> PMID: 32020859
50. Davis T, Doyle H, Tobias V, Ellison DW, Ziegler DS. Case report of spontaneous resolution of a congenital glioblastoma. *Pediatrics.* 2016;137(4):e20151241. <https://doi.org/10.1542/peds.2015-1241> PMID: 26980879
51. Siveen KS, Uddin S, Mohammad RM. Targeting acute myeloid leukemia stem cell signaling by natural products. *Mol Cancer.* 2017;16(1):13. <https://doi.org/10.1186/s12943-016-0571-x> PMID: 28137265
52. Van Etten RA. Aberrant cytokine signaling in leukemia. *Oncogene.* 2007;26(47):6738–49. <https://doi.org/10.1038/sj.onc.1210758> PMID: 17934482
53. Jimenez-Pascual A, Siebzehnrubl FA. Fibroblast growth factor receptor functions in glioblastoma. *Cells.* 2019;8(7):715. <https://doi.org/10.3390/cells8070715> PMID: 31337028
54. Bruntz RC, Lindsley CW, Brown HA. Phospholipase D signaling pathways and phosphatidic acid as therapeutic targets in cancer. *Pharmacol Rev.* 2014;66(4):1033–79. <https://doi.org/10.1124/pr.114.009217> PMID: 25244928
55. Jalnapurkar SS, Pawar AS, George SS, Antony C, Somers P, Grana J, et al. PHF6 suppresses self-renewal of leukemic stem cells in AML. *Leukemia.* 2024;38(9):1938–48. <https://doi.org/10.1038/s41375-024-02340-5> PMID: 39004675
56. Hurtado R, Guirales F, Tirado CA. ASXL1 gene in AML. *J Assoc Genet Technol.* 2021;47(2):60–8. PMID: 34140438
57. Nepstad I, Hatfield KJ, Grønningsæter IS, Reikvam H. The PI3K-Akt-mTOR signaling pathway in human acute myeloid leukemia (AML) cells. *Int J Mol Sci.* 2020;21(8):2907. <https://doi.org/10.3390/ijms21082907> PMID: 32326335
58. Modarres P, Mohamadi Farsani F, Nekouie AA, Vallian S. Meta-analysis of gene signatures and key pathways indicates suppression of JNK pathway as a regulator of chemo-resistance in AML. *Sci Rep.* 2021;11(1):12485. <https://doi.org/10.1038/s41598-021-91864-2> PMID: 34127725
59. Li X, Wu C, Chen N, Gu H, Yen A, Cao L, et al. PI3K/Akt/mTOR signaling pathway and targeted therapy for glioblastoma. *Oncotarget.* 2016;7(22):33440–50. <https://doi.org/10.18632/oncotarget.7961> PMID: 26967052
60. Solomon DA, Kim J-S, Cronin JC, Sibenaller Z, Ryken T, Rosenberg SA, et al. Mutational inactivation of PTPRD in glioblastoma multiforme and malignant melanoma. *Cancer Res.* 2008;68(24):10300–6. <https://doi.org/10.1158/0008-5472.CAN-08-3272> PMID: 19074898
61. Kaburagi T, Yamato G, Shiba N, Yoshida K, Hara Y, Tabuchi K, et al. Clinical significance of RAS pathway alterations in pediatric acute myeloid leukemia. *Haematologica.* 2022;107(3):583–92. <https://doi.org/10.3324/haematol.2020.269431> PMID: 33730843
62. Zafar N, Ghias K, Fadoo Z. Genetic aberrations involved in relapse of pediatric acute myeloid leukemia: a literature review. *Asia Pac J Clin Oncol.* 2021;17(5):e135–41. <https://doi.org/10.1111/ajco.13367> PMID: 32573082
63. Advani AS, Hunger SP, Burnett AK. Acute leukemia in adolescents and young adults. *Semin Oncol.* 2009;36(3):213–26. <https://doi.org/10.1053/j.seminoncol.2009.03.007> PMID: 19460579
64. Rudat S, Pfaus A, Cheng YY, Holtmann J, Ellegast JM, Bühler C, et al. RET-mediated autophagy suppression as targetable co-dependence in acute myeloid leukemia. *Leukemia.* 2018;32(10):2189–202. <https://doi.org/10.1038/s41375-018-0102-4> PMID: 29654265

65. Laneve P, Rea J, Caffarelli E. Long noncoding RNAs: emerging players in medulloblastoma. *Frontiers in Pediatrics*. 2019;7:67.
66. Dunham C. Uncommon pediatric tumors of the posterior fossa: pathologic and molecular features. *Childs Nerv Syst*. 2015;31(10):1729–37. <https://doi.org/10.1007/s00381-015-2735-1> PMID: 26351226
67. Tsubota S, Kadomatsu K. Origin and initiation mechanisms of neuroblastoma. *Cell Tissue Res*. 2018;372(2):211–21. <https://doi.org/10.1007/s00441-018-2796-z> PMID: 29445860
68. Su L, Guo W, Lou L, Nie S, Zhang Q, Liu Y, et al. EGFR-ERK pathway regulates CSN6 to contribute to PD-L1 expression in glioblastoma. *Mol Carcinog*. 2020;59(5):520–32. <https://doi.org/10.1002/mc.23176> PMID: 32134157
69. Deng D, Yang S, Wang X. Long non-coding RNA SNHG16 regulates cell behaviors through miR-542-3p/HNF4 α axis via RAS/RAF/MEK/ERK signaling pathway in pediatric neuroblastoma cells. *Biosci Rep*. 2020;40(5):BSR20200723. <https://doi.org/10.1042/BSR20200723> PMID: 32412051
70. Li N, Wu P, Wang Z, Shen Y, Zhang L, Xue F, et al. Smoking-related cancer death among men and women in an ageing society (China 2020-2040): a population-based modelling study. *Tob Control*. 2023;32(2):163–9. <https://doi.org/10.1136/tobaccocontrol-2020-056444> PMID: 34725269
71. Moody S, Senkin S, Islam SMA, Wang J, Nasrollahzadeh D, Cortez Cardoso Penha R, et al. Mutational signatures in esophageal squamous cell carcinoma from eight countries with varying incidence. *Nat Genet*. 2021;53(11):1553–63. <https://doi.org/10.1038/s41588-021-00928-6> PMID: 34663923
72. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non-small-cell lung cancer. *N Engl J Med*. 2017;376(22):2109–21. <https://doi.org/10.1056/NEJMoa1616288> PMID: 28445112
73. Mo L, Wei B, Liang R, Yang Z, Xie S, Wu S, et al. Exploring potential biomarkers for lung adenocarcinoma using LC-MS/MS metabolomics. *J Int Med Res*. 2020;48(4):300060519897215. <https://doi.org/10.1177/0300060519897215> PMID: 32316791
74. Zhang T, Joubert P, Ansari-Pour N, Zhao W, Hoang PH, Lokanga R, et al. Genomic and evolutionary classification of lung cancer in never smokers. *Nat Genet*. 2021;53(9):1348–59. <https://doi.org/10.1038/s41588-021-00920-0> PMID: 34493867
75. Ceraudo E, Horioka M, Mattheisen JM, Hitchman TD, Moore AR, Kazmi MA, et al. Direct evidence that the GPCR CysLTR2 mutant causative of uveal melanoma is constitutively active with highly biased signaling. *J Biol Chem*. 2021;296:100163. <https://doi.org/10.1074/jbc.RA120.015352> PMID: 33288675
76. Desrichard A, Kuo F, Chowell D, Lee K-W, Riaz N, Wong RJ, et al. Tobacco smoking-associated alterations in the immune microenvironment of squamous cell carcinomas. *J Natl Cancer Inst*. 2018;110(12):1386–92. <https://doi.org/10.1093/jnci/djy060> PMID: 29659925
77. Viegas C, Pereira DS, Fonte P. Insights into nanomedicine for head and neck cancer diagnosis and treatment. *Materials*. 2022;15(6):2086.
78. Dong Y, Lu S, Wang Z, Liu L. CCTs as new biomarkers for the prognosis of head and neck squamous cancer. *Open Med (Wars)*. 2020;15(1):672–88. <https://doi.org/10.1515/med-2020-0114> PMID: 33313411
79. Zhou K, Chen J, Wu J, Xu Y, Wu Q, Yue J, et al. Profilin 2 promotes proliferation and metastasis of head and neck cancer cells by regulating PI3K/AKT/ β -catenin signaling pathway. *Oncol Res*. 2019;27(9):1079–88. <https://doi.org/10.3727/096504019X15579146061957> PMID: 31122311
80. Marziliano A, Teckie S, Diefenbach MA. Alcohol-related head and neck cancer: summary of the literature. *Head Neck*. 2020;42(4):732–8. <https://doi.org/10.1002/hed.26023> PMID: 31777131
81. Lovinger DM. Presynaptic ethanol actions: potential roles in ethanol seeking. *Handb Exp Pharmacol*. 2018;248:29–54. https://doi.org/10.1007/164_2017_76 PMID: 29204712
82. Mulder DJ, Pooni A, Mak N, Hurlbut DJ, Basta S, Justinich CJ. Antigen presentation and MHC class II expression by human esophageal epithelial cells: role in eosinophilic esophagitis. *Am J Pathol*. 2011;178(2):744–53. <https://doi.org/10.1016/j.ajpath.2010.10.027> PMID: 21281807
83. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K, et al. Body Fatness and Cancer—Viewpoint of the IARC Working Group. *N Engl J Med*. 2016;375(8):794–8. <https://doi.org/10.1056/NEJMs1606602> PMID: 27557308
84. Vucenik I, Stains JP. Obesity and cancer risk: evidence, mechanisms, and recommendations. *Ann N Y Acad Sci*. 2012;1271(1):37–43. <https://doi.org/10.1111/j.1749-6632.2012.06750.x> PMID: 23050962
85. Tu T, Bühler S, Bartenschlager R. Chronic viral hepatitis and its association with liver cancer. *Biol Chem*. 2017;398(8):817–37. <https://doi.org/10.1515/hsz-2017-0118> PMID: 28455951

86. Crosbie J, Varma J, Mansfield J. *Yersinia enterocolitica* infection in a patient with hemochromatosis masquerading as proximal colon cancer with liver metastases: report of a case. *Dis Colon Rectum*. 2005;48(2):390–2. <https://doi.org/10.1007/s10350-004-0777-7> PMID: 15812589
87. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*. 2022;50(W1):W216–W221.