Pan-cancer proteogenomic landscape of whole-genome doubling reveals putative therapeutic targets in various cancer types

Author names and affiliations

Eunhyong Chang[1,2], Hee Sang Hwang[3], Kyu Jin Song[4,5], Kwoneel Kim[6,7], Min-Sik Kim[8,9,10], Se Jin Jang[3,11], Kwang Pyo Kim[4,5], Sungyong You[12,13], Joon-Yong An[1,2,14,*]

[1] Department of Integrated Biomedical and Life Science, Korea University, Seoul 02841, Republic of Korea

[2] L-HOPE Program for Community-Based Total Learning Health Systems, Korea University, Seoul 02841, Republic of Korea

[3] Department of Pathology, University of Ulsan College of Medicine, Asan Medical Center, Seoul 05505, Republic of Korea

[4] Department of Applied Chemistry, Institute of Natural Science, Kyung Hee University, Yongin 17104, Republic of Korea

[5] Department of Biomedical Science and Technology, Kyung Hee Medical Science Research Institute, Kyung Hee University, Seoul 02453, Republic of Korea

[6] Department of Biology, Kyung Hee University, Seoul 02447, Republic of Korea

[7] Department of Biomedical and Pharmaceutical Sciences, Kyung Hee University, Seoul 02447, Republic of Korea

[8] Department of New Biology, DGIST, Daegu,42988, Republic of Korea

[9] New Biology Research Center, DGIST, Daegu 42988, Republic of Korea

[10] Center for Cell Fate Reprogramming and Control, DGIST, Daegu 42988, Republic of Korea

[11] Oncoclew Co., Ltd., 31, Ttukseom-ro 1-gil, Seongdong-gu, Seoul 04778, Republic of Korea

[12] Department of Urology, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

[13] Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

[14] School of Biosystem and Biomedical Science, College of Health Science, Korea University, Seoul 02841, Republic of Korea

***Correspondence:** joonan30@korea.ac.kr (Joon-Yong An)

Abstract

Background: Whole-genome doubling (WGD) is prevalent in cancer and drives tumor development and chromosomal instability. Driver mutations in mitotic cell cycle genes and cell cycle upregulation have been reported as the major molecular underpinnings of WGD tumors. However, the underlying genomic signatures and regulatory networks involved in gene transcription and kinase phosphorylation remain unclear. Here, we aimed to comprehensively decipher the molecular landscape underlying WGD tumors.

Methods: We performed a pan-cancer proteogenomic analysis and compared 10 cancer types by integrating genomic, transcriptomic, proteomic, and phosphoproteomic datasets from the Clinical Proteomic Tumor Analysis Consortium (CPTAC). We also integrated the cancer dependency data of each cancer cell line and the survival properties of each cancer patient to propose promising therapeutic targets for patients with WGD.

Results: Our study delineated distinct copy number signatures characterizing WGD-positive tumors into three major groups: highly unstable genome, focal instability, and tetraploidy. Furthermore, the analysis revealed the heterogeneous mechanisms underlying WGD across cancer types with specific structural variation patterns. Upregulation of the cell cycle and downregulation of the immune response were found to be specific to certain WGD tumor types. Transcription factors (TFs) and kinases exhibit cancer-specific activities, emphasizing the need for tailored therapeutic approaches.

Conclusion: This study introduces an integrative approach to identify potential TF targets for drug development, highlighting BPTF as a promising candidate for the treatment of head and neck squamous cell carcinoma. Additionally, drug repurposing strategies have been proposed, suggesting potential drugs for the treatment of WGD-associated cancers. Our findings offer insights into the heterogeneity of WGD and have implications for precision medicine approaches for cancer treatment.

Keywords: Whole-genome doubling, Pan-cancer, Multi-omics, Cancer genomics, Personalized medicine, Genome instability, BPTF, Head and neck squamous cell carcinoma

Background

Whole-genome doubling (WGD) is prevalent across cancer subtypes, promoting tumor development and generating chromosomal instability (CIN). WGD plays an important role in tumorigenesis by cushioning the deleterious mutations and rapidly accumulating genetic abnormalities. However, WGD is also associated with tolerance to genomic instability, which leads to cell death [1-3]. Previous studies have investigated the genomic alterations associated with WGD, revealing that *TP53* mutations and defects in the E2F-mediated G1 arrest are common in WGD-positive tumors [1]. Moreover, gene expression studies have highlighted the enrichment of genes involved in cellular proliferation, mitotic spindle formation, and DNA repair, whereas inflammatory pathways are downregulated in WGD-positive tumors [2]. These findings depict the overall pan-cancer characteristics of WGD; however, the role of WGD can be highly heterogeneous [2,4-7].

Recent proteomic studies have delineated novel mechanisms underlying diverse cancer subtypes [8]. These investigations have revealed associations between certain multiomic subtypes that are indicative of WGD. For example, in head and neck squamous cell carcinoma (HNSCC), a distinct molecular subtype has been characterized by high CIN and upregulated cell cycle pathways at both the proteome and phosphoproteome levels [9]. Similar WGD-associated subtypes have been detected in non-small cell lung cancer (NSCLC), endometrial cancer, breast cancer (BRCA), colon cancer, and glioblastoma (GBM) [10-14]. Despite mounting evidence implicating WGD-related subtypes across multiple cancers, a pan-cancer multiomics investigation focusing on WGD remains elusive. In addition, the proteomic features and kinase activities governing WGD in cancer are yet to be elucidated. Furthermore, the development of therapeutic strategies that specifically target WGD-positive tumors remains an unmet need.

In this study, we aimed to conduct a pan-cancer proteogenomic analysis to delineate the genomic and proteomic landscapes of WGD across ten types of cancer by integrating genomic, transcriptomic, proteomic, and phosphoproteomic data sets. Our objective was to characterize the molecular pathways, transcription factor (TF) regulation, and kinase phosphorylation networks enriched in association with the WGD. Finally, we explored the potential drug targets and repositioning strategies for patients with WGDs.

Results

Pan-cancer analysis to identify CN signatures underlying WGD

We sought to explore the proteogenomic features associated with WGD by analyzing comprehensive genomic, transcriptomic, proteomic, and phosphoproteomic datasets obtained from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) (**Figure 1A**). The dataset comprised 1,060 patients representing 10 types of cancer: breast cancer (BRCA) [12], clear cell renal cell carcinoma (CCRCC) [15], colon adenocarcinoma (COAD) [13], glioblastoma (GBM) [14], high-grade serous carcinoma (HGSC) [16,17], head and neck squamous cell carcinoma (HNSCC) [9], lung adenocarcinoma (LUAD) [11], lung squamous cell carcinoma (LSCC) [10], pancreatic ductal adenocarcinoma (PDAC) [18], and uterine corpus endometrial carcinoma (UCEC) [12]. By determining the WGD status of each sample, we observed a bimodal distribution of patients, indicating the existence of two distinct groups of cancers irrespective of the cancer type (**Figure 1B**). Consistent with previous estimates, approximately 42% of the tumors (440 of 1,060 samples) exhibited at least 1 occurrence of WGD during their evolutionary process[1,2]. We also identified substantial variability in the occurrence of WGD across different tumor types, with HGSC showing the highest prevalence (83%, 65/78 samples) and PDAC showing the lowest (9.4%, 13/139 samples) (**Figure 1C**).

Mutation signature analyses demonstrated that WGD was associated with specific copy number signatures (**Table S1B**). Based on the 25 CN signature values of the WGD-positive samples, different cancer types exhibited varying combinations of CN signatures, suggesting that the underlying mechanisms of WGD might be distinct across cancer types (**Figure 1D**). In patients with LSCC, those with WGD showed significant enrichment for CN7 (FDR = 3.11×10^{-5}), indicating chromothripsis amplification, as well as for CN15 (FDR = 2.87×10^{-2}), signifying chromosomal loss-of-heterozygosity (LOH) with twice-genome-doubling (**Figure S1A; Table S1B**). Similarly, patients with HNSCC showed significant enrichment of CN15 in WGD-positive samples (FDR = 4.63×10^{-6}). Patients with LUAD showed enrichment of the CN7 signature (FDR = 5.39×10^{-4}). WGD in BRCA and HGSC correlated significantly with CN11, a signature of focal LOH, with two WGD events (BRCA, FDR=1.71×10^{-5}; HGSC, FDR=2.68×10^{-13}), suggesting that WGD in these malignancies occurred within a focally unstable genomic context. A majority of WGD samples from CCRCC, COAD, GBM, PDAC, and UCEC were found to be significantly enriched for CN2 (FDR < 0.05), indicating tetraploidy. Based on these observations, we defined three distinct WGD status in LSCC,

LUAD, and HNSCC as "WGD type 1," that in BRCA and HGSC as "WGD type 2," and that in CCRCC, COAD, GBM, PDAC, and UCEC as "WGD type 3." The WGD status was present in a coherent grouping of samples, wherein cancer types with similar CN signatures exhibited spatial proximity (**Figure 1E**).

We further evaluated the driver mutations associated with WGD in each tumor type. Despite previous reports suggesting an enrichment of *TP53* mutations in WGD tumors across diverse cancers [1,2], we identified a significant enrichment of *TP53* mutations exclusively in the WGD-positive samples of BRCA, COAD, UCEC, and LUAD ($p < 0.05$; Fisher's exact test) (**Figure 1F; Table S1C**). Although no common gene mutations were identified, we observed that over 100 mutations were significantly enriched in association with WGD in LSCC, LUAD, and HNSCC (WGD type 1), whereas other cancer types exhibited fewer than 15 significant gene mutations associated with WGD. As a highly unstable genome has been linked to a higher tumor mutational burden (TMB) [19,20], we speculated that the TMB would be higher in WGD type 1. We found significantly higher TMB in WGD-positive samples than in WGD-negative samples across 10 cancer types ($p = 3.12 \times 10^{-4}$; Wilcoxon rank-sum test) (**Figure S1A**), and in LSCC, LUAD, HNSCC, and BRCA when compared across individual cancer types ($p < 0.05$; Wilcoxon rank-sum test) (**Figure S1B**). WGD-positive samples in COAD exhibited significantly lower TMB than WGD-negative samples ($p = 2.13 \times 10^{-4}$; Wilcoxon rank-sum test). Nevertheless, *TP53* and *APC* mutations remained significantly enriched in samples with WGD in COAD (*TP53*, $p = 1.75 \times 10^{-3}$; *APC*, $p = 2.31 \times 10^{-2}$; Fisher's exact test) (**Figure 1F**). This is consistent with the findings of prior investigations suggesting an association between *APC* mutations and aneuploidy in COAD [21-23]. These observations imply that genomic instability is a catalyst for WGD in LSCC, LUAD, and HNSCC, whereas *TP53* and *APC* mutations may serve as primary drivers of WGD in COAD. In summary, our findings revealed distinct CN signatures of WGD, allowing us to define three types of WGD.

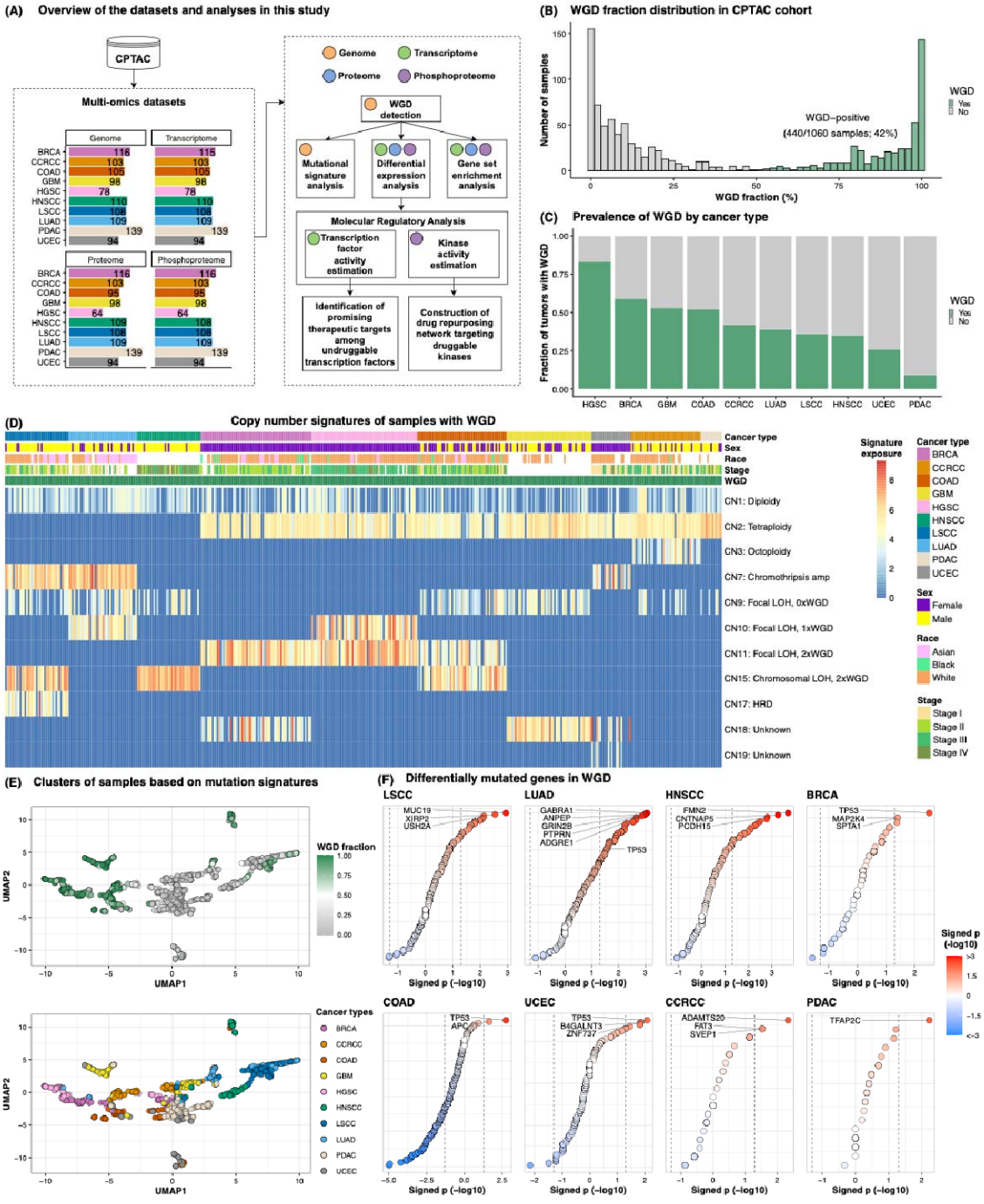

Figure 1. Copy number signatures associated with WGD

(A) Schematic representation of the multi-omics datasets collected for each cancer type and the subsequent analysis pipeline. BRCA, breast cancer; CCRCC, Clear cell renal cell carcinoma; COAD, colon adenocarcinoma; GBM, Glioblastoma; HGSC, High-grade serous carcinoma; HNSCC, Head and neck squamous cell carcinoma; LSCC, Lung squamous cell carcinoma; LUAD, Lung adenocarcinoma; PDAC, Pancreatic ductal adenocarcinoma; UCEC, Uterine corpus endometrial carcinoma. **(B)** Distribution of the WGD fraction within the CPTAC cohort, displaying a bimodal pattern. **(C)** Prevalence of WGD by cancer type. **(D)** Heatmap based on copy number signature exposure values in samples with WGD. Copy number signatures were

derived from the Catalogue Of Somatic Mutations In Cancer (COSMIC). **(E)** UMAP plot based on signature exposure values of single-base-substitution, double-base-substitution, indel, and copy number alteration. Each dot indicates samples and is colored based on WGD fraction (top) and cancer types (bottom). **(F)** Dot plot depicting differentially mutated genes in WGD in each cancer type. The position of each dot along the x-axis and the color of the dots indicate signed p-values obtained from Fisher's exact test. Only the top three significant genes were labeled in each cancer type, except for COAD and PDAC, for which only two and one significant genes, respectively, were noted. In LUAD, *PTPRN, GRIN2B, ANPEP*, and *ADGRE1* genes exhibited the same p-values, and *TP53* was additionally labeled.

Enrichment of distinct pathways in WGD in each cancer type

Previous studies have reported activation of the cell cycle pathway and inactivation of the immune response pathway in tumors with WGD [1,2,24,25]. Given the various CN signatures underlying WGD across cancer types, we conducted a sample-level pathway enrichment test with 1,060 samples and examined the biological pathways enriched in WGD (**Figure 2A**). Overall, WGD-positive tumors were significantly affected by several pathways (**Table S2A**), which were subsequently categorized into four major pathway groups: cell motility, immune response, cell cycle, and metabolism. WGD type 1 tumors, characterized by a highly unstable genome, showed significant upregulation of the cell cycle and downregulation of immune response pathways (**Figure 2B and 2C**), which is consistent with previous reports [1,2,24,25]. In contrast, WGD type 2 tumors showed significant enrichment in the dTTP metabolism pathway, which is responsible for DNA synthesis and maintenance [26,27] ($p < 0.01$; Wilcoxon rank-sum test), and the DNA endoreduplication pathway, a known mechanism inducing WGD [28,29] ($p < 0.05$; Wilcoxon rank-sum test) (**Figure 2SA and 2SB**). Among other cancer types, WGD tumors in COAD showed significant activation of the Wnt signaling pathway, possibly attributed to *APC* mutation, in line with the findings of previous studies [30-32] ($p = 2.09 \times 10^{-3}$; Wilcoxon rank-sum test) (**Figure 2SC**). These results emphasize that upregulation of the cell cycle and downregulation of the immune response are specific to WGD type 1 tumors, suggesting diverse functional attributes of WGD across distinct cancer types.

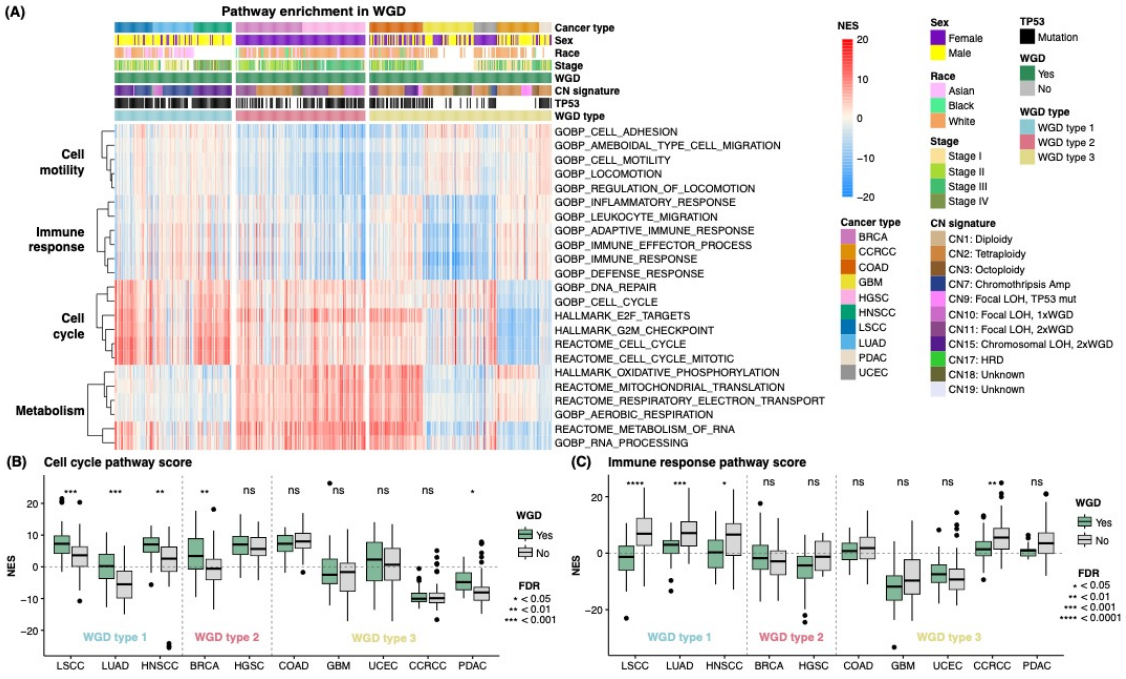

Figure 2. Pathway enrichment in WGD-positive tumors

(A) Normalized enrichment scores (NES) of pathways related to cell motility, immune response, cell cycle, and metabolism in WGD-positive tumors. Pathways exhibiting significance (FDR < 0.05) and a log fold-change greater than 0.25 are shown. **(B–C)** Boxplot comparing NES score of cell cycle pathway (HALLMARK_E2F_TARGETS) and immune response pathway (GOBP_IMMUNE_RESPONSE) between WGD-positive and WGD-negative tumors in individual cancer types.

WGD-specific TFs as potential therapeutic targets

To identify potential therapeutic targets for treating WGD across various tumor types, we first estimated TF activity using the TF-target gene interaction network database [33] and the gene expression levels of target genes (see Methods). Out of the 1,134 TFs analyzed, the E2F family and MYC TFs showed significant activation (FDR < 0.05) in pan-cancer WGD tumors, indicating the role of cell cycle regulation in WGD pathophysiology [34,35] (**Figure 3A; Table S3A**). In contrast, the TFs that were significantly downregulated in the WGD-positive tumors were predominantly associated with immune response pathways. While different TFs were activated in WGD-positive tumors across different cancer types (**Figure 3B**), E2F1, E2F2, E2F3, E2F4, and MYC were common TFs with significant activation in five cancer types (LSCC, LUAD, HNSCC, BRCA, and PDAC) that showed upregulation of the cell cycle pathway (**Figure 2B**).

Since many TFs have been deemed as 'undruggable' due to their structural complexities and a lack of tractable binding sites [36,37], we introduced an integrative

framework to prioritize WGD-activated TFs as putative therapeutic targets (**Figures 3C and 3SA–I; Table S3C**). Among the 99 TFs that exhibited significant activation in each tumor type (FDR < 0.1), we first selected TFs that were detected in CPTAC proteomics data with high confidence (FDR < 0.01). We then filtered TFs exhibiting significantly upregulated protein expression in WGD-positive tumors compared to WGD-negative tumors. We found that the protein expression of IRX2, GLI2, E2F4, and TFDP1 in LSCC, E2F3 in LUAD, BPTF, REST, and SFPQ in HNSCC, and SREBF2 (also known as SREBP2) in BRCA was significantly upregulated (FDR < 0.05, integrated hypothesis test) (**Figure 3D**). The cancer dependencies of these TFs in each cancer cell line were then evaluated by comparing the dependencies between cells with and without WGD. Among the nine TFs that showed significant upregulation at the protein level, four TFs including IRX2 in LSCC and BPTF, REST, and SFPQ in HNSCC exhibited significantly decreased viability in cells with WGD upon CRISPR-mediated depletion ($p < 0.1$, Wilcoxon rank-sum test) (**Figure 3E**). Finally, we selected the TFs with high protein expression levels that were associated with poor prognosis. We found a significant association between high BPTF protein expression and unfavorable prognosis in HNSCC ($p = 0.017$, log-rank test) (**Figure 3F**). BPTF is reported as a co-factor of c-MYC leading to c-MYC-driven proliferation and G1 to S progression [38], and we also observed significant activation of MYC in HNSCC (FDR = 5.66×10^{-3}). Our findings suggest that deactivating BPTF in patients with WGD-positive HNSCC could slow down cancer cell proliferation, which may ultimately benefit patient survival.

Several TFs did not meet the criteria outlined in our integrative framework; however, they remain potential candidates for further consideration as therapeutic targets. In LSCC, GLI2 showed both significant activation (FDR = 9.37×10^{-2}) and upregulation at both mRNA and protein levels (mRNA, FDR = 1.71×10^{-8}, Wald test; protein, FDR = 1.72×10^{-3}, integrated hypothesis test) (**Figure S3B; Table S3C**). As GLI2 is known to promote cell proliferation and cancer cell survival by upregulating the expression of antiapoptotic proteins in LSCC [39], degrading GLI2 may be a promising strategy for the treatment of patients with LSCC harboring WGD. Additionally, TFDP1 and TFDP2, which are partner proteins of the E2F family crucial for the G1 to S phase transition [40,41], were identified as potential therapeutic targets for WGD-positive samples in LSCC, HNSCC, PDAC, and HGSC, in which these TFs were significantly activated (FDR < 0.1) (**Figure S3A, S3B, S3E, and S3I**). In LUAD, E2F3 showed significant activation (FDR = 3.60×10^{-4}) and upregulation at both mRNA and protein levels

(mRNA, FDR = 3.16×10^{-5}, Wald test; protein, FDR = 1.83×10^{-3}, integrated hypothesis test) (**Figure S3C**). Therefore, E2F3 inhibitors such as Edifoligide could be a potential treatment for LUAD in patients with WGD. Furthermore, in BRCA, SREBF2 showed significant activation (FDR = 8.67×10^{-2}) along with upregulated protein expression (FDR = 2.42×10^{-2}, integrated hypothesis test) (**Figure S3D**). Targeting SREBF2 could be effective for WGD-positive BRCA samples since SREBF2 is known to regulate the synthesis of cholesterol which is crucial for cancer cell viability in breast cancer cells [42-44]. Taken together, our integrative framework incorporated protein expression, cancer dependency, and association with patient survival to delineate effective therapeutic targets among significantly activated TFs. Based on our analysis, we propose that BPTF could potentially serve as a therapeutic target for WGD-positive samples in HNSCC, thereby improving patient prognosis.

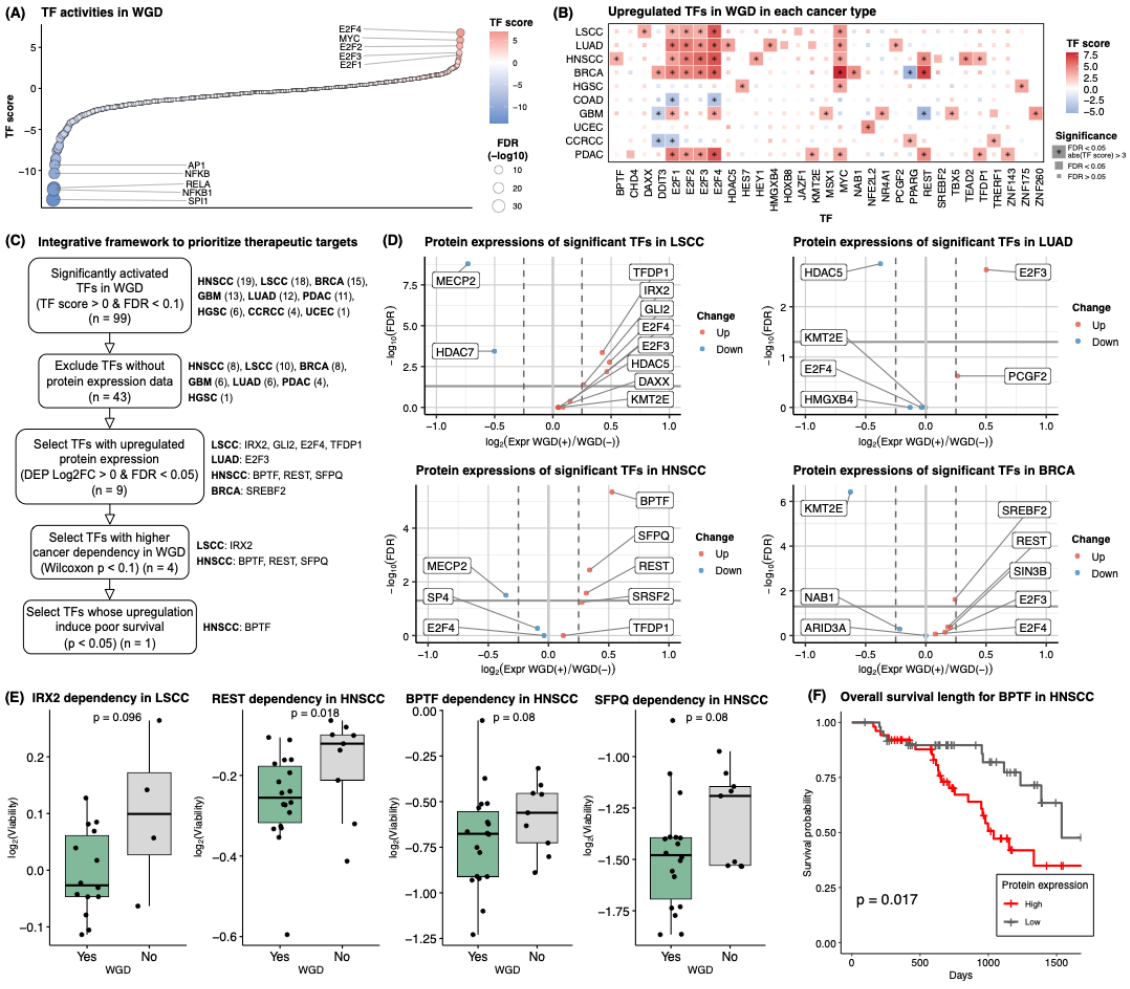

Figure 3. Deciphering potential therapeutic targets among activated TFs in WGD

(A) Dot plot depicting estimated TF scores in WGD-positive versus WGD-negative tumors across pan-cancer analysis. The top and bottom five TFs, ranked by TF score, are labeled. **(B)**

Significantly upregulated TFs in WGD in each tumor type. The tiles are colored based on the TF estimation score, and TFs meeting the criteria of FDR < 0.05 and TF score > 3 are denoted with asterisks. **(C)** Schematic diagram showing the filtering processes to identify effective therapeutic targets. The number of TFs or the name of TFs remaining after each step is indicated on the right. **(D)** Volcano plots of differentially expressed proteins among significant TFs in WGD-positive tumors compared to WGD-negative tumors. **(E)** Boxplots comparing $\log_2$ cell viability between WGD-positive and WGD-negative cells after depleting genes with CRISPR in each cancer cell line. **(F)** Kaplan–Meier curve of overall survival based on protein expression levels of BPTF in HNSCC.

Drug repurposing strategies to target key kinases in WGD

As kinases are one of the well-known druggable proteins [37,45], we next questioned whether we can suggest treatment strategies by deciphering kinase activities in WGD and matching appropriate drugs targeting these kinases. To estimate kinase activities in the WGD in each tumor type, we utilized a kinase-substrate interaction database [46] and phosphoprotein expression data of substrates (see Methods). Consistently, we observed that cancer types with upregulated cell cycle pathways exhibited CDK1 and CDK2 activation (FDR < 0.1) (**Figure 4A; Table S4A**). The protein expression of these cyclin-dependent kinases was also significantly upregulated (FDR < 0.05, integrated hypothesis test) (**Figure 4B**). Apart from cell cycle regulation, kinases involved in the DNA damage response and hypoxia-induced autophagy were significantly activated in some tumors. CSNK2A1, which phosphorylates key components of DNA damage and repair pathways [47], was significantly activated in LSCC and LUAD (LSCC, FDR = 7.62×10^{-5}; LUAD, FDR = 1.05×10^{-3}) (**Figure 4A; Table S4A**). Additionally, in LSCC, PAK4 showed significant activation in WGD (FDR = 6.70×10^{-3}). It is known to promote proliferation and suppress apoptosis [48,49], and its overexpression has been linked to poor prognosis in NSCLC [50]. In BRCA, both PRKAA2 (also known as AMPK2) and ULK1 were significantly activated (PRKAA2, FDR = 8.34×10^{-2}; ULK1, FDR = 3.21×10^{-2}). PRKAA2 induces autophagy during glucose starvation by phosphorylating ULK1 [51,52]. These results imply that although WGD can drive tumorigenesis by accelerating cell proliferation, it can also induce DNA damage and hypoxia, necessitating kinase activation to inhibit apoptosis, as previously noted [1-3].

Next, we sought to identify putative drug targets for kinases with upregulated expression in WGD using protein-protein and protein-drug interactions [53] (see Methods). Our analysis revealed that nintedanib was the most promising treatment for WGD in LSCC

and LUAD, as it interacts with a majority of significantly activated kinases in these cancers (**Figure 4C; Table S4B**). Nintedanib is a small-molecule tyrosine kinase inhibitor that is used for NSCLC patients along with docetaxel after the first line of chemotherapy [54]. For WGD in COAD, bosutinib and ruxolitinib were suggested to target PLK1 kinase, a pivotal regulator of mitotic events frequently overexpressed in colon cancers [55-58]. These drugs have been reported to induce apoptosis in colon cancer cells [59,60]. Sunitinib was recommended as the most suitable drug for BRCA patients with WGD, supported by its efficacy in BRCA reported in some studies [61,62], although caution is advised when applying it to metastatic breast cancer patients [63]. Overall, our findings unraveled cancer-type-specific kinase activations in WGD contributing to cell proliferation, DNA damage response, and hypoxia-driven autophagy. Based on these findings, we proposed FDA-approved drugs for treating patients with WGD in each tumor type.

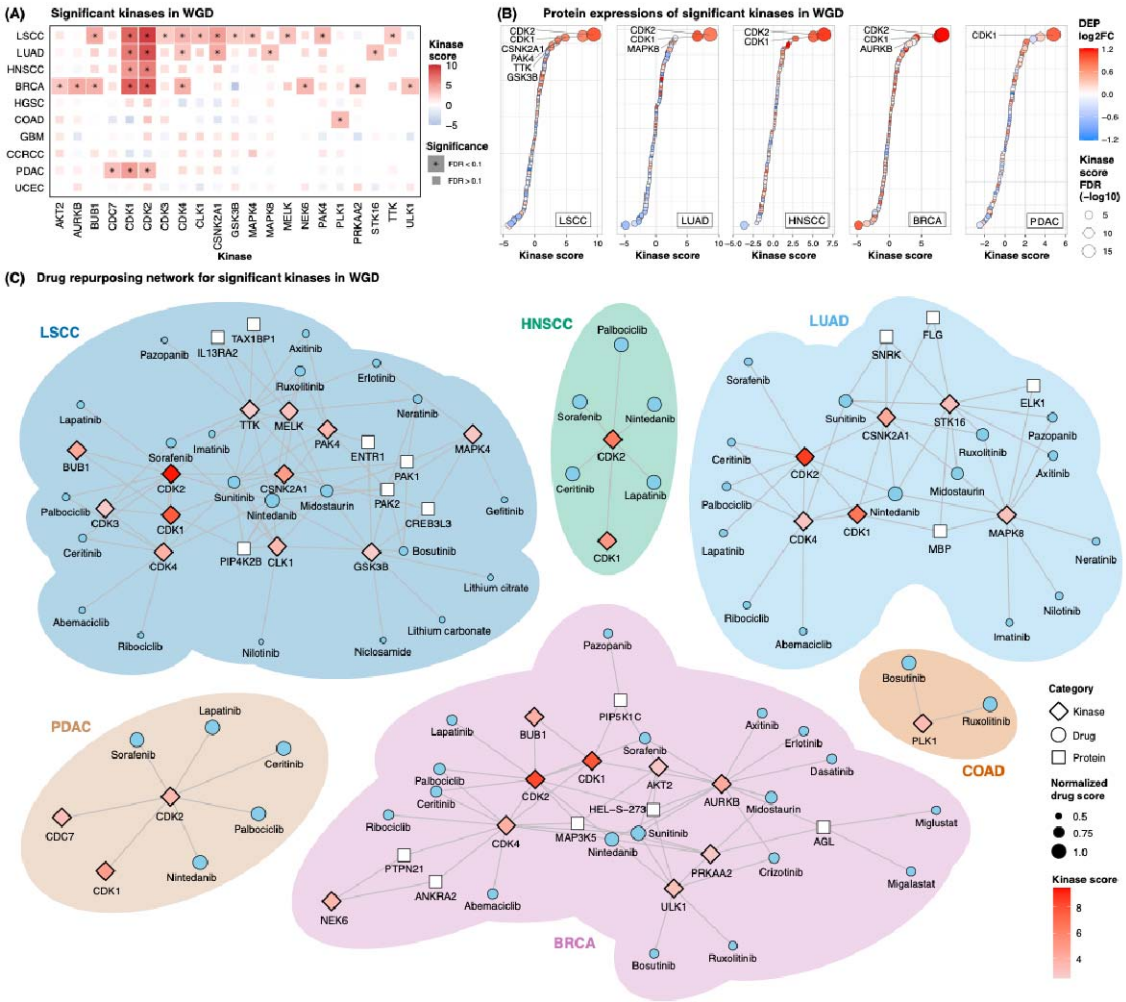

Figure 4. Drug repurposing strategies to target key kinases in WGD

(A) Kinases with significant upregulation of expression in association with WGD in each tumor type. The tiles are colored based on the kinase estimation score, and kinases showing significant upregulation are denoted with asterisks. **(B)** Dot plot depicting differentially expressed proteins among significantly upregulated kinases in WGD-positive tumors compared to WGD-negative tumors. Kinases with significant upregulation of protein expression (FDR < 0.05 and log2-fold-change > 0), as well as activation (FDR < 0.1), have been labeled. **(C)** Drug repurposing network for key kinases associated with WGD in each tumor type.

Discussion

Our study elucidated the proteogenomic characteristics of WGD in a cancer-type-specific manner. CN signatures were used to characterize WGD tumors into three major groups: WGD type 1, with a highly unstable genome (LSCC, LUAD, and HNSCC); type 2, with focal instability (BRCA and HGSC); and type 3, with tetraploidy (CCRCC, COAD, GBM, PDAC, and UCEC). This classification seems to align with cancer-specific profiles of structural variation [64]. Triple-negative breast cancer and ovarian cancer are enriched in an SV class, characterized by a high burden of deletions, duplications, and templated insertion chains. LSCC and HNSCC are characterized by an enriched breakage fusion bridge cycle, which is a mechanism of chromosomal instability [65]. The presence of CN7 and CN15 signatures underscores the occurrence of WGD in a genomically unstable environment at the chromosomal level [66].

While the etiology of CN18 remains unknown, we observed its enrichment in WGD-positive BRCA and GBM. CN18 has been linked to TP53 mutation in BRCA [66], and our study demonstrates a significant enrichment of TP53 mutation in WGD-positive BRCA samples. This implies CN18 in BRCA may be a WGD signature driven by TP53 mutation. Additionally, CN18 has been associated with hypoxia, which is known to induce polyploid giant cancer cells with stem-like phenotypes in GBM [67,68]. As a hypoxic condition in GBM has been linked to advanced tumor stage and invasion [69], CN18 in GBM may imply hypoxia-driven WGD with an invasive cell state.

In recent studies, WGD tumors have been predominantly characterized by an upregulated cell cycle and a downregulated immune response [1,2,24,25]. However, our study demonstrated that this pattern is specific to WGD type 1 (LSCC, LUAD, and HNSCC) and does not entirely represent other types of WGD-positive tumors. For these cancer types, significant activation of E2F and MYC seemed to regulate the cell cycle pathway in WGD-

positive tumors. Instead, the activities of TFs and kinases were more cancer-type-specific than consistent across WGD types.

Our study has important therapeutic implications for WGD tumors. Given the challenges of targeting undruggable TFs [36,37,45], we introduced an integrative approach to identify putative TF targets for new drug development, requiring high protein expression levels, increased cancer dependency, and poorer prognosis in patients with WGD. Based on these criteria, we identified BPTF as a promising target for treating patients with HNSCC harboring WGD. For kinases associated with WGD, we employed a drug repurposing strategy using search algorithms to identify the most suitable drugs targeting these kinases. Drug repositioning is a cost-effective approach for identifying effective treatments without the need for extensive time or resources for new drug development [70]. We found nintedanib for LSCC and LUAD, bosutinib and ruxolitinib for COAD, and sunitinib for BRCA were recommended as potential drugs for treating WGD patients. Further studies are warranted to evaluate the efficacy of drugs targeting WGD-specific TFs and of repurposed drugs in patient cohorts with WGD. This additional investigation is essential for advancing precision medicine approaches tailored to WGD-associated cancers.

Despite our comprehensive analysis, there are limitations within our study. For cancer types included in CPTAC phase 2 (BRCA, COAD, and HGSC), whole-exome sequencing data were utilized to infer WGD, a method potentially less precise than whole-genome sequencing. In addition, our research covered only 1,060 samples across 10 tumor types, indicating the necessity for larger-scale proteogenomic investigations to fully understand WGD characteristics in various cancers. Future studies should consider employing single-cell and spatial proteomics to gain insights into the cellular dynamics and properties of WGD tumor cells. Furthermore, the effectiveness of BPTF inhibitors in WGD-positive HNSCC, among other drugs for WGD-positive tumors, requires further validation. Despite these limitations, our study on the proteogenomic landscape of WGD across cancers could provide therapeutic insights for treating WGD-related subtypes in individual tumor types, supporting the groundwork for future research endeavors.

Acknowledgments

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2019M3E5D3073568 to Joon-Yong An) and by the Ministry of Science and ICT (NRF-2022R1A2C2013377 to Min-Sik Kim), the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare, Republic of Korea (RS-2023-00304839 to Kwoneel Kim), and Korea University to Joon-Yong An. Eunhyong Chang received a scholarship from the Kwanjeong Educational Foundation and the Brain Korea (BK21) FOUR education program.

Conflict of interest

Kwang Pyo Kim is the CEO of NioBiopharmaceuticals, Inc. Se Jin Jang is the chief executive officer of Oncoclew, Co., Ltd. All other authors report no conflict of interest.

Data availability statement

Public CPTAC datasets are available within 11 publications [9-18]. Genome and transcriptome data were downloaded from the GDC data portal (https://portal.gdc.cancer.gov). Global proteomic and phosphoproteomic data were downloaded from LinkedOmics (https://www.linkedomics.org/). All data generated during this study are included in Supporting Information.

Authors' contributions

Eunhyong Chang and Joon-Yong An designed the study; Eunhyong Chang analyzed the data; Eunhyong Chang, Hee Sang Hwang, Kyu Jin Song, Kwoneel Kim, Min-Sik Kim, Se Jin Jang, Kwang Pyo Kim, Sungyong You, Joon-Yong An wrote and reviewed the manuscript; all authors approved the final version of the manuscript.

List of abbreviations: Whole-genome doubling (WGD), Chromosomal instability (CIN), Breast cancer (BRCA), Clear cell renal cell carcinoma (CCRCC), Colon adenocarcinoma (COAD), Glioblastoma (GBM), High-grade serous carcinoma (HGSC), Head and neck squamous cell carcinoma (HNSCC), Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LSCC), Pancreatic ductal adenocarcinoma (PDAC), Uterine corpus endometrial carcinoma (UCEC) Loss-of-heterozygosity (LOH), Tumor mutational burden (TMB), Copy number (CN), Transcription factor (TF)

406 **Methods**

407 **Data collection and preprocessing**

408 We obtained genomic, transcriptomic, global proteomic, and phosphoproteomic data from
409 the Clinical Proteomic Tumor Analysis Consortium. The mutation annotation format file for
410 each sample and the segment-level copy number variant (CNV) text file for the samples in
411 CPTAC phase-3 (comprising clear cell renal cell carcinoma [CCRCC], GBM, HNSCC, lung
412 squamous cell carcinoma [LSCC], lung adenocarcinoma [LUAD], pancreatic ductal
413 adenocarcinoma [PDAC], and uterine corpus endometrial carcinoma [UCEC]) were
414 downloaded from the GDC data portal (https://portal.gdc.cancer.gov). Transcriptomic data
415 were downloaded from the GDC data portal. Global proteomic and phosphoproteomic data
416 were downloaded from LinkedOmics (https://www.linkedomics.org/).

417 To standardize the global phosphoprotein data across various cancer types, we
418 initially normalized the data using z-scores. Subsequently, global proteins and
419 phosphoproteins with > 30% missing values in each cancer-type sample were excluded,
420 followed by k-nearest neighbor imputation with k=5 to address missing values.

421

422 **CNV calling and WGD detection**

423 As segment-level CN data for BRCA, COAD, and HGSC were unavailable, we acquired BAM
424 files from the GDC data portal and performed the CNV calling process. FACETS v0.16.0 [71] was
425 employed to discern allele-specific CN information. The input for FACETS consisted of paired
426 tumor-normal BAM files and a VCF file containing common and germline polymorphic sites
427 downloaded from https://www.ncbi.nlm.nih.gov/variation/docs/human_variation_vcf/.

428 Using major CN and minor CN data derived from segment-level CNV data, we defined
429 samples as "WGD-positive" if over 50% of their autosomal genome exhibited a major CN
430 greater than or equal to two [1].

431

432 **Mutational signature analysis**

433 We conducted a mutational signature analysis using the COSMIC signature database v3 [72]
434 and R package Sigminer v2.1.5 [73]. Non-negative matrix factorization was employed to
435 determine the number of signature groups or factorization ranks. This involved creating a
436 tumor-by-component matrix with 50 runs and checking the ranks ranging from 2 to 12. Each
437 signature was identified using the COSMIC signature with the highest cosine similarity.

Subsequently, hierarchical clustering was performed, and the samples were assigned to one of the signatures based on the consensus matrix.

Analysis of differential expression

To discern the differentially expressed genes between WGD-positive and WGD-negative samples, we employed the "DESeq" function from the DESeq2 R package [74]. The raw count data were used for the generation of a DESeqDataSet using the "DESeqDataSetFromMatrix" function. Following a variance stabilizing transformation and the exclusion of data points with a mean raw count less than 50, we conducted the differential expression analysis utilizing the "DESeq" function, which is grounded in the negative binomial distribution. The DEG analysis was performed using the following formula:

$$design = \sim WGD.status$$

To detect the differential expression of global and phospho-proteins in the WGD samples, we employed an integrated hypothesis testing method following the methodology proposed by Hwang et al [75]. Briefly, we performed t-tests, median difference tests, and Wilcoxon tests and combined the p-values from the statistical tests using Stouffer's method. This approach enhances the accuracy of identifying true-positive data elements compared with traditional methods when applied to biological data.

Gene set enrichment analysis

We conducted a gene set enrichment analysis to elucidate the activated biological pathways in each sample. The "zscore" function from the R package GSVA [76] was utilized for analyzing transcriptome, global proteome, and phospho-proteome data. Pathways from various databases, including hallmark ("h.all.v2023.1.Hs.symbols.gmt.txt"), KEGG ("c2.cp.kegg.v2023.1.Hs.symbols.gmt.txt"), Reactome ("c2.cp.reactome.v2023.1.Hs.symbols.gmt.txt"), GO ("c5.go.v2023.1.Hs.symbols.gmt.txt"), and Wikipathway ("c2.cp.wikipathways.v2023.1.Hs.symbols.gmt.txt") sourced from MSigDB were incorporated into the gene set variation analysis. Subsequently, we performed a t-test between WGD-positive tumors and WGD-negative tumors to identify significantly regulated pathways in the context of WGD.

Estimation of transcription factor and kinase activity

We used the data corresponding to transcription factor (TF)-target interactions from CollecTRI [77] to estimate TF activity. The log-2-fold change values derived from the analysis of differentially expressed genes (DEGs) served as inputs for the identification of TF activity in WGD-positive samples. For estimating kinase activity, we employed kinase-substrate interactions sourced from OmniPathR [46]. The log-2-fold change values obtained from the analysis of differentially expressed phosphorylated proteins were used as inputs to identify the kinase activity in WGD-positive samples. Subsequently, we employed the "run_mlm" function within the decoupleR R package [33] to infer both TF and kinase activities, with the minimum size of regulons set to 1.

Cancer dependency and druggability analysis

For the comparison of cancer dependencies between WGD-positive and WGD-negative cells, we downloaded "CRISPR (DepMap Public 23Q2+Score, Chronos)," "RNAi (Achilles+DRIVE+Marcotte, DEMETER2)," and "Aneuploidy" data for each cancer cell line from DepMap (https://depmap.org/portal/). We subset the cell lines with annotated genome doubling statuses. Cell lines with one or more instances of genome doubling were classified as WGD-positive, whereas those without genome doubling were categorized as WGD-negative. We conducted a one-sided Wilcoxon rank-sum test to assess whether the WGD-positive cells exhibited lower viability than the WGD-negative cells. To evaluate the druggability of each TF, we used the DGIdb (v5.0.5) [78]. TFs lacking interactions with drugs were deemed "undruggable," those with interactions with FDA-approved drugs were designated as "FDA-approved," and those interacting with known drugs but not with FDA-approved ones were labeled as "druggable."

Survival analysis

The Kaplan–Meier estimation model was used to perform survival analysis. The survival duration of the patients and death events were used as inputs for the analysis. To identify the TFs associated with poor prognosis, we compared survival curves between samples with the top 50% and bottom 50% protein expression for each significantly activated TF in WGD (false discovery rate [FDR] < 0.1).

Drug repurposing network analysis

DrugSt. One (v1.2.0) [53] was used to build the drug repurposing network. Significantly upregulated kinases (FDR < 0.1) in the WGD for each tumor type were used as inputs for the search algorithms. NeDRex (v2.21.0) was used for protein-protein and protein-drug interaction searches. For the protein-protein interaction network, we used a multi-Steiner algorithm, setting the number of trees to five, tolerance to five, and hub penalty to 0.5. For the protein-drug interaction network, we used a harmonic centrality algorithm, setting the hub penalty to 0.5 and the result size to 50 and excluding indirect and non-approved drugs.

Supplemental tables

Table S1.xlsx: Sample overview and copy number signatures associated with WGD

Table S2.xlsx: Enriched pathways in WGD

Table S3.xlsx: Activated transcription factor in WGD and their druggability

Table S4.xlsx: Activated kinases in WGD and drug repurposing network

Supplemental figures

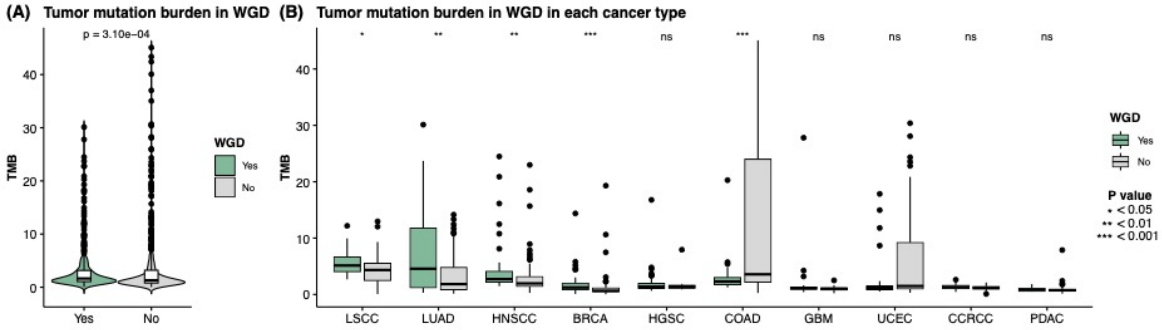

Figure S1. Tumor mutation burden in WGD

(A) Violin plot comparing tumor mutation burden between WGD-positive and WGD-negative tumors across 10 cancer types. **(B)** Boxplots comparing tumor mutation burden between WGD-positive and WGD-negative tumors within each cancer type.

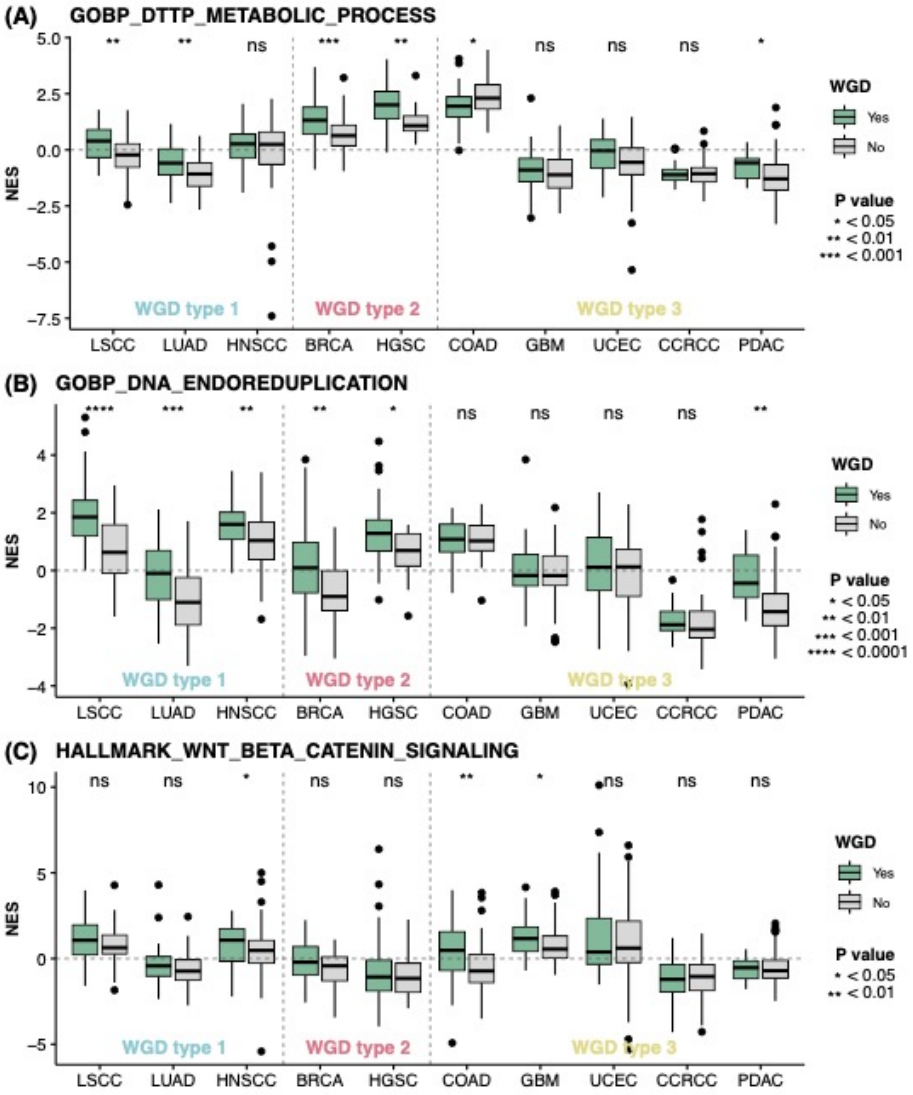

Figure S2. Pathway enrichment in WGD-positive tumors

(A-C) Boxplot comparing Normalized enrichment scores (NES) score of dTTP metabolism pathway (GOBP_DTTP_METABOLIC_PROCESS), DNA endoreduplication pathway (GOBP_DNA_ENDOREDUPLICATION), and Wnt signaling pathway (HALLMARK_WNT_BETA_CATENIN_SIGNALING) between WGD-positive and WGD-negative tumors in individual cancer types.

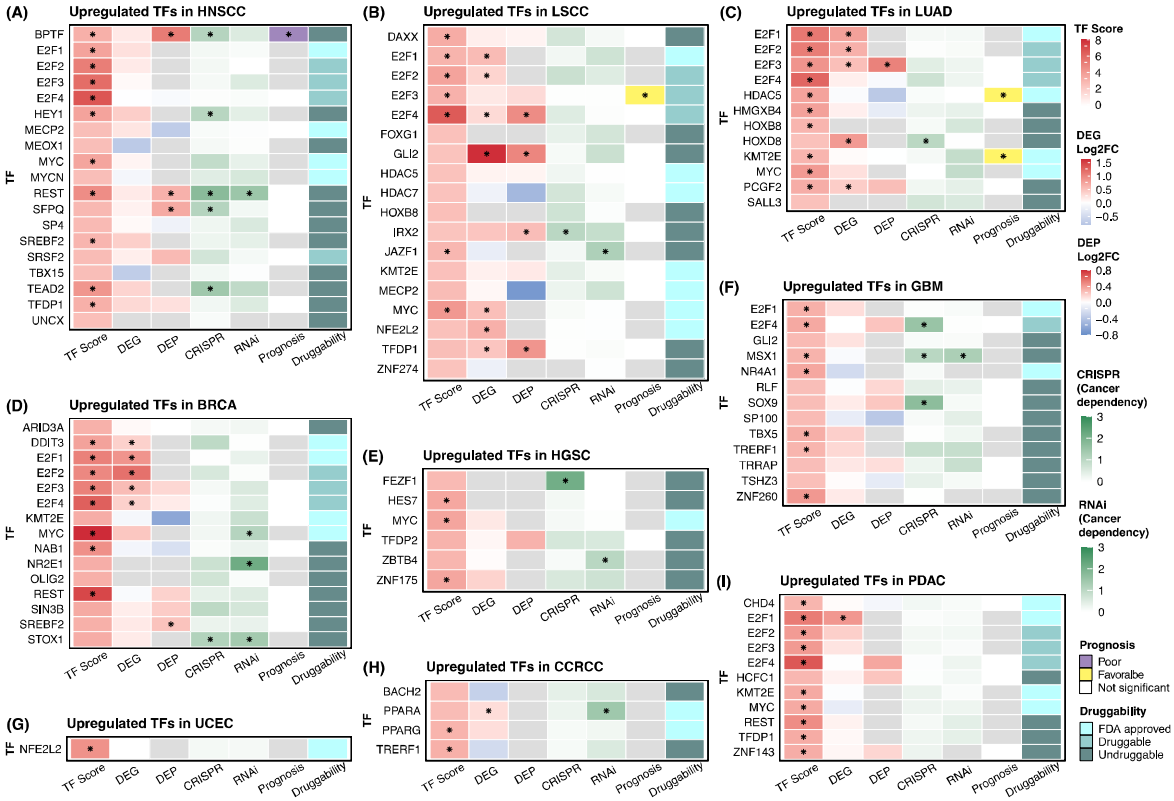

Figure S3. Activated TFs in WGD in each cancer type

(A-I) Significantly upregulated TFs in WGD in each cancer type are shown. Significant features are indicated as an asterisk (TF score, FDR < 0.05; DEG, FDR < 0.05; DEP, FDR < 0.05; CRISPR, p < 0.1; RNAi, p < 0.1; Prognosis, p <0.05).

References

1. Bielski CM, Zehir A, Penson AV, et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet*. Aug 2018;50(8):1189-1195. doi:10.1038/s41588-018-0165-1
2. Quinton RJ, DiDomizio A, Vittoria MA, et al. Whole-genome doubling confers unique genetic vulnerabilities on tumour cells. *Nature*. Feb 2021;590(7846):492-497. doi:10.1038/s41586-020-03133-3
3. Zack TI, Schumacher SE, Carter SL, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet*. Oct 2013;45(10):1134-40. doi:10.1038/ng.2760
4. Lopez S, Lim EL, Horswell S, et al. Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat Genet*. Mar 2020;52(3):283-293. doi:10.1038/s41588-020-0584-7
5. Ganem NJ, Cornils H, Chiu S-Y, et al. Cytokinesis failure triggers hippo tumor suppressor pathway activation. *Cell*. 2014;158(4):833-848.
6. Senovilla L, Vitale I, Martins I, et al. An immunosurveillance mechanism controls cancer cell ploidy. *Science*. 2012;337(6102):1678-1684.
7. Chang E, An J-Y. Whole-genome doubling is a double-edged sword: the heterogeneous role of whole-genome doubling in various cancer types. *BMB reports*. 2024:6157-6157.
8. Heo YJ, Hwa C, Lee GH, Park JM, An JY. Integrative Multi-Omics Approaches in Cancer Research: From Biological Networks to Clinical Subtypes. *Mol Cells*. Jul 31 2021;44(7):433-443. doi:10.14348/molcells.2021.0042
9. Huang C, Chen L, Savage SR, et al. Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell*. Mar 8 2021;39(3):361-379 e16. doi:10.1016/j.ccell.2020.12.007
10. Satpathy S, Krug K, Jean Beltran PM, et al. A proteogenomic portrait of lung squamous cell carcinoma. *Cell*. Aug 5 2021;184(16):4348-4371 e40. doi:10.1016/j.cell.2021.07.016
11. Gillette MA, Satpathy S, Cao S, et al. Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. *Cell*. Jul 9 2020;182(1):200-225 e35. doi:10.1016/j.cell.2020.06.013
12. Krug K, Jaehnig EJ, Satpathy S, et al. Proteogenomic Landscape of Breast Cancer Tumorigenesis and Targeted Therapy. *Cell*. Nov 25 2020;183(5):1436-1456 e31. doi:10.1016/j.cell.2020.10.036
13. Vasaikar S, Huang C, Wang X, et al. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell*. May 2 2019;177(4):1035-1049 e19. doi:10.1016/j.cell.2019.03.030
14. Wang LB, Karpova A, Gritsenko MA, et al. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell*. Apr 12 2021;39(4):509-528 e20. doi:10.1016/j.ccell.2021.01.006
15. Clark DJ, Dhanasekaran SM, Petralia F, et al. Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell*. Oct 31 2019;179(4):964-983 e31. doi:10.1016/j.cell.2019.10.007
16. McDermott JE, Arshad OA, Petyuk VA, et al. Proteogenomic Characterization of Ovarian HGSC Implicates Mitotic Kinases, Replication Stress in Observed Chromosomal Instability. *Cell Rep Med*. Apr 21 2020;1(1)doi:10.1016/j.xcrm.2020.100004

17. Hu Y, Pan J, Shah P, et al. Integrated Proteomic and Glycoproteomic Characterization of Human High-Grade Serous Ovarian Carcinoma. *Cell Rep*. Oct 20 2020;33(3):108276. doi:10.1016/j.celrep.2020.108276
18. Cao L, Huang C, Cui Zhou D, et al. Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell*. Sep 16 2021;184(19):5031-5052 e26. doi:10.1016/j.cell.2021.08.023
19. Cai L, Bai H, Duan J, et al. Epigenetic alterations are associated with tumor mutation burden in non-small cell lung cancer. *J Immunother Cancer*. Jul 26 2019;7(1):198. doi:10.1186/s40425-019-0660-7
20. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. Apr 19 2017;9(1):34. doi:10.1186/s13073-017-0424-2
21. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *cell*. 1990;61(5):759-767.
22. Fodde R, Smits R, Clevers H. APC, signal transduction and genetic instability in colorectal cancer. *Nature reviews cancer*. 2001;1(1):55-67.
23. Powell SM, Zilz N, Beazer-Barclay Y, et al. APC mutations occur early during colorectal tumorigenesis. *Nature*. 1992;359(6392):235-237.
24. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*. Jan 20 2017;355(6322)doi:10.1126/science.aaf8399
25. Cohen-Sharir Y, McFarland JM, Abdusamad M, et al. Aneuploidy renders cancer cells vulnerable to mitotic checkpoint inhibition. *Nature*. Feb 2021;590(7846):486-491. doi:10.1038/s41586-020-03114-6
26. Nordlund P, Reichard P. Ribonucleotide reductases. *Annu Rev Biochem*. 2006;75:681-706. doi:10.1146/annurev.biochem.75.103004.142443
27. Chabes A, Georgieva B, Domkin V, Zhao X, Rothstein R, Thelander L. Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase. *Cell*. 2003;112(3):391-401.
28. Davoli T, Denchi EL, de Lange T. Persistent telomere damage induces bypass of mitosis and tetraploidy. *Cell*. Apr 2 2010;141(1):81-93. doi:10.1016/j.cell.2010.01.031
29. Zeng J, Hills SA, Ozono E, Diffley JFX. Cyclin E-induced replicative stress drives p53-dependent whole-genome duplication. *Cell*. Feb 2 2023;186(3):528-542 e14. doi:10.1016/j.cell.2022.12.036
30. Rubinfeld B, Souza B, Albert I, et al. Association of the APC gene product with β-catenin. *Science*. 1993;262(5140):1731-1734.
31. Su L-K, Vogelstein B, Kinzler KW. Association of the APC tumor suppressor protein with catenins. *Science*. 1993;262(5140):1734-1737.
32. Korinek V, Barker N, Morin PJ, et al. Constitutive transcriptional activation by a β-catenin-Tcf complex in APC−/− colon carcinoma. *Science*. 1997;275(5307):1784-1787.
33. Badia-i-Mompel P, Vélez Santiago J, Braunger J, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances*. 2022;2(1):vbac016.
34. Dyson N. The regulation of E2F by pRB-family proteins. *Genes & development*. 1998;12(15):2245-2262.
35. Bretones G, Delgado MD, Leon J. Myc and cell cycle control. *Biochim Biophys Acta*. May 2015;1849(5):506-16. doi:10.1016/j.bbagrm.2014.03.013

631 36. Henley MJ, Koehler AN. Advances in targeting 'undruggable' transcription factors
632 with small molecules. *Nat Rev Drug Discov*. Sep 2021;20(9):669-688. doi:10.1038/s41573-
633 021-00199-0
634 37. Xie X, Yu T, Li X, et al. Recent advances in targeting the "undruggable" proteins: from
635 drug discovery to clinical trials. *Signal Transduct Target Ther*. Sep 6 2023;8(1):335.
636 doi:10.1038/s41392-023-01589-z
637 38. Richart L, Carrillo-de Santa Pau E, Rio-Machin A, et al. BPTF is required for c-MYC
638 transcriptional activity and in vivo tumorigenesis. *Nat Commun*. Jan 5 2016;7:10153.
639 doi:10.1038/ncomms10153
640 39. Huang L, Walter V, Hayes DN, Onaitis M. Hedgehog-GLI signaling inhibition
641 suppresses tumor growth in squamous lung cancer. *Clin Cancer Res*. Mar 15
642 2014;20(6):1566-75. doi:10.1158/1078-0432.CCR-13-2195
643 40. Girling R, Partridge JF, Bandara LR, et al. A new component of the transcription factor
644 DRTF1/E2F. *Nature*. 1993;362(6415):83-87.
645 41. Lam EW-F, La Thangue NB. DP and E2F proteins: coordinating transcription with cell
646 cycle progression. *Current opinion in cell biology*. 1994;6(6):859-866.
647 42. Goldstein JL, DeBose-Boyd RA, Brown MS. Protein sensors for membrane sterols. *Cell*.
648 Jan 13 2006;124(1):35-46. doi:10.1016/j.cell.2005.12.022
649 43. Bengoechea-Alonso MT, Ericsson J. SREBP in signal transduction: cholesterol
650 metabolism and beyond. *Curr Opin Cell Biol*. Apr 2007;19(2):215-22.
651 doi:10.1016/j.ceb.2007.02.004
652 44. Griffiths B, Lewis CA, Bensaad K, et al. Sterol regulatory element binding protein-
653 dependent regulation of lipid synthesis supports cell survival and tumor growth. *Cancer &*
654 *metabolism*. 2013;1:1-21.
655 45. Peng F, Liao M, Qin R, et al. Regulated cell death (RCD) in cancer: key pathways and
656 targeted therapies. *Signal Transduct Target Ther*. Aug 13 2022;7(1):286. doi:10.1038/s41392-
657 022-01110-y
658 46. Turei D, Korcsmaros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for
659 literature-curated signaling pathway resources. *Nat Methods*. Nov 29 2016;13(12):966-967.
660 doi:10.1038/nmeth.4077
661 47. Rabalski AJ, Gyenis L, Litchfield DW. Molecular Pathways: Emergence of Protein
662 Kinase CK2 (CSNK2) as a Potential Target to Inhibit Survival and DNA Damage Response and
663 Repair Pathways in Cancer Cells. *Clin Cancer Res*. Jun 15 2016;22(12):2840-7.
664 doi:10.1158/1078-0432.CCR-15-1314
665 48. Callow MG, Clairvoyant F, Zhu S, et al. Requirement for PAK4 in the anchorage-
666 independent growth of human cancer cell lines. *J Biol Chem*. Jan 4 2002;277(1):550-8.
667 doi:10.1074/jbc.M105732200
668 49. Li X, Minden A. PAK4 functions in tumor necrosis factor (TNF) alpha-induced survival
669 pathways by facilitating TRADD binding to the TNF receptor. *J Biol Chem*. Dec 16
670 2005;280(50):41192-200. doi:10.1074/jbc.M506884200
671 50. Cai S, Ye Z, Wang X, et al. Overexpression of P21-activated kinase 4 is associated with
672 poor prognosis in non-small cell lung cancer and promotes migration and invasion. *J Exp Clin*
673 *Cancer Res*. May 15 2015;34(1):48. doi:10.1186/s13046-015-0165-2
674 51. Kim J, Kundu M, Viollet B, Guan KL. AMPK and mTOR regulate autophagy through
675 direct phosphorylation of Ulk1. *Nat Cell Biol*. Feb 2011;13(2):132-41. doi:10.1038/ncb2152

52. Egan DF, Shackelford DB, Mihaylova MM, et al. Phosphorylation of ULK1 (hATG1) by AMP-activated protein kinase connects energy sensing to mitophagy. *Science*. 2011;331(6016):456-461.
53. Maier A, Hartung M, Abovsky M, et al. Drugst. One-A plug-and-play solution for online systems medicine and network-based drug repurposing. *Arxiv*. 2023;
54. Roth GJ, Binder R, Colbatzky F, et al. Nintedanib: from discovery to the clinic. *J Med Chem*. Feb 12 2015;58(3):1053-63. doi:10.1021/jm501562a
55. Roshak AK, Capper EA, Imburgia C, Fornwald J, Scott G, Marshall LA. The human polo-like kinase, PLK, regulates cdc2/cyclin B through phosphorylation and activation of the cdc25C phosphatase. *Cellular signalling*. 2000;12(6):405-411.
56. Dai W, Wang Q, Traganos F. Polo-like kinases and centrosome regulation. *Oncogene*. Sep 9 2002;21(40):6195-200. doi:10.1038/sj.onc.1205710
57. Takahashi T, Sano B, Nagata T, et al. Polo-like kinase 1 (PLK1) is overexpressed in primary colorectal cancers. *Cancer Sci*. Feb 2003;94(2):148-52. doi:10.1111/j.1349-7006.2003.tb01411.x
58. Han D-p, Zhu Q-l, Cui J-t, et al. Polo-like kinase 1 is overexpressed in colorectal cancer and participates in the migration and invasion of colorectal cancer cells. *Medical science monitor: international medical journal of experimental and clinical research*. 2012;18(6):BR237.
59. Mologni L, Cleris L, Magistroni V, et al. Valproic acid enhances bosutinib cytotoxicity in colon cancer cells. *Int J Cancer*. Apr 15 2009;124(8):1990-6. doi:10.1002/ijc.24158
60. Li X, Wang Z, Zhang S, Yao Q, Chen W, Liu F. Ruxolitinib induces apoptosis of human colorectal cancer cells by downregulating the JAK1/2-STAT1-Mcl-1 axis. *Oncol Lett*. May 2021;21(5):352. doi:10.3892/ol.2021.12613
61. Korashy HM, Maayah ZH, Al Anazi FE, et al. Sunitinib Inhibits Breast Cancer Cell Proliferation by Inducing Apoptosis, Cell-cycle Arrest and DNA Repair While Inhibiting NF-kappaB Signaling Pathways. *Anticancer Res*. Sep 2017;37(9):4899-4909. doi:10.21873/anticanres.11899
62. Burstein HJ, Elias AD, Rugo HS, et al. Phase II study of sunitinib malate, an oral multitargeted tyrosine kinase inhibitor, in patients with metastatic breast cancer previously treated with an anthracycline and a taxane. *J Clin Oncol*. Apr 10 2008;26(11):1810-6. doi:10.1200/JCO.2007.14.5375
63. Wang D, Xiao F, Feng Z, et al. Sunitinib facilitates metastatic breast cancer spreading by inducing endothelial cell senescence. *Breast Cancer Res*. Sep 29 2020;22(1):103. doi:10.1186/s13058-020-01346-y
64. Hadi K, Yao X, Behr JM, et al. Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs. *Cell*. Oct 1 2020;183(1):197-210 e32. doi:10.1016/j.cell.2020.08.006
65. McClintock B. The behavior in successive nuclear divisions of a chromosome broken at meiosis. *Proceedings of the National Academy of Sciences*. 1939;25(8):405-416.
66. Steele CD, Abbasi A, Islam SMA, et al. Signatures of copy number alterations in human cancer. *Nature*. Jun 2022;606(7916):984-991. doi:10.1038/s41586-022-04738-6
67. Qu Y, Zhang L, Rong Z, He T, Zhang S. Number of glioma polyploid giant cancer cells (PGCCs) associated with vasculogenic mimicry formation and tumor grade in human glioma. *Journal of Experimental & Clinical Cancer Research*. 2013;32:1-7.
68. Heddleston JM, Li Z, McLendon RE, Hjelmeland AB, Rich JN. The hypoxic microenvironment maintains glioblastoma stem cells and promotes reprogramming towards

723 a cancer stem cell phenotype. *Cell Cycle*. Oct 15 2009;8(20):3274-84.
724 doi:10.4161/cc.8.20.9701
725 69. Xie Q, Mittal S, Berens ME. Targeting adaptive glioblastoma: an overview of
726 proliferation and invasion. *Neuro Oncol*. Dec 2014;16(12):1575-84.
727 doi:10.1093/neuonc/nou147
728 70. Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and
729 recommendations. *Nat Rev Drug Discov*. Jan 2019;18(1):41-58. doi:10.1038/nrd.2018.168
730 71. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity
731 analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res*. Sep 19
732 2016;44(16):e131. doi:10.1093/nar/gkw520
733 72. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In
734 Cancer. *Nucleic Acids Res*. Jan 8 2019;47(D1):D941-D947. doi:10.1093/nar/gky1015
735 73. Wang S, Li H, Song M, et al. Copy number signature analysis tool and its application in
736 prostate cancer reveals distinct mutational processes and clinical outcomes. *PLoS Genet*.
737 May 2021;17(5):e1009557. doi:10.1371/journal.pgen.1009557
738 74. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
739 RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8
740 75. Hwang D, Rust AG, Ramsey S, et al. A data integration methodology for systems
741 biology. *Proceedings of the National Academy of Sciences*. 2005;102(48):17296-17301.
742 76. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray
743 and RNA-seq data. *BMC bioinformatics*. 2013;14:1-15.
744 77. Muller-Dott S, Tsirvouli E, Vazquez M, et al. Expanding the coverage of regulons from
745 high-confidence prior knowledge for accurate estimation of transcription factor activities.
746 *Nucleic Acids Res*. Nov 10 2023;51(20):10934-10949. doi:10.1093/nar/gkad841
747 78. Cannon M, Stevenson J, Stahl K, et al. DGIdb 5.0: rebuilding the drug-gene interaction
748 database for precision medicine and drug discovery platforms. *Nucleic Acids Res*. Jan 5
749 2024;52(D1):D1227-D1235. doi:10.1093/nar/gkad1040
750