

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

Dear Editor,

Whole-genome doubling (WGD) occurs in various cancer types and plays a crucial role in tumour development and genomic instability.^{1,2} However, the proteogenomic characteristics and molecular regulators governing WGD have yet to be elucidated. By integrating large-scale multi-omics data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC),³ we classified three types of WGD tumours across cancer types. This study elucidates the mutational signatures, molecular pathways, transcription factor (TF) regulation and kinase phosphorylation networks enriched in tumours with WGD to explore potential drug targets.

Our study integrated genomic, transcriptomic, proteomic and phosphoproteomic data from 1060 samples representing 10 cancer types including 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 squamous cell carcinoma (LSCC), lung adenocarcinoma (LUAD), pancreatic ductal adenocarcinoma (PDAC) and uterine corpus endometrial carcinoma (UCEC) (Figure 1A). WGD was found to be prevalent in 42% of tumours, with the highest in HGSC (83%) and the lowest in PDAC (9.4%) (Figure 1B,C; Table S1A). Although no clinical phenotypes were associated with WGD status across pan-cancer, the advanced tumour stage was linked to WGD status in HNSCC (false discovery rate [FDR] < .05) (Figure S1A,B).

Mutation signature analyses showed that WGD was associated with specific copy number (CN) signatures (Figure 1D; Table S1B). WGD in LSCC, LUAD and HNSCC showed enrichment for either or both CN7 and CN15 signatures (FDR < .05). Since CN7 indicates chromothripsis amplification and CN15 indicates chromosomal loss-of-heterozygosity (LOH) with twice-genome-doubling,⁴ WGD in LSCC, LUAD and HNSCC may have occurred within a highly unstable genome. WGD in BRCA and

HGSC was enriched for CN11 (focal LOH with twice-genome-doubling) (FDR < .05), suggesting that WGD in these malignancies occurred within a focally unstable genome. Many WGD samples from CCRCC, COAD, GBM, PDAC and UCEC were found to be enriched for CN2 (FDR < .05), indicating tetraploidy. Consequently, three WGD types were defined: 'WGD type 1 (highly unstable genome)' in LSCC, LUAD and HNSCC; 'WGD type 2 (focal instability)' in BRCA and HGSC; 'WGD type 3 (tetraploidy)' in CCRCC, COAD, GBM, PDAC and UCEC (Figure 1E).

Driver mutations associated with WGD were identified in each tumour type. Whereas previous studies suggest enrichment of *TP53* mutations in WGD tumours,^{1,2} we identified enrichment of *TP53* mutations exclusively in the WGD-positive samples of BRCA, COAD, UCEC and LUAD ($p < .05$) (Figure 1F; Table S1D). Moreover, we found that more than 80 gene amplifications resulted in gene and protein expression upregulation (involved in cell cycle and metabolism) in WGD-positive LSCC, LUAD and HNSCC (WGD type 1) (Figure S1C; Table S1D,E). Additionally, more than 100 mutations were enriched in WGD type 1 ($p < .05$), whereas other cancer types exhibited <15 significant gene mutations associated with WGD. As a highly unstable genome has a high level of tumour mutational burden (TMB),⁵ WGD type 1 had a higher TMB in WGD-positive than in WGD-negative tumours ($p < .05$; Wilcoxon rank-sum test) (Figure S1D,E). COAD exhibited lower TMB in WGD-positive tumours than in WGD-negative tumours ($p < .001$; Wilcoxon rank-sum test) and *TP53* and *APC* mutations were augmented in WGD-positive COAD ($p < .05$; Fisher's exact test) (Figure 1F). Genomic instability is therefore a catalyst for WGD in LSCC, LUAD and HNSCC, whereas *TP53* and *APC* mutations may serve as primary drivers of WGD in COAD.

Pathway analysis revealed distinct biological processes enriched in WGD tumours across different cancer types (Figure 2A; Table S2A,B). Although previous studies have

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Clinical and Translational Medicine* published by John Wiley & Sons Australia, Ltd on behalf of Shanghai Institute of Clinical Bioinformatics.

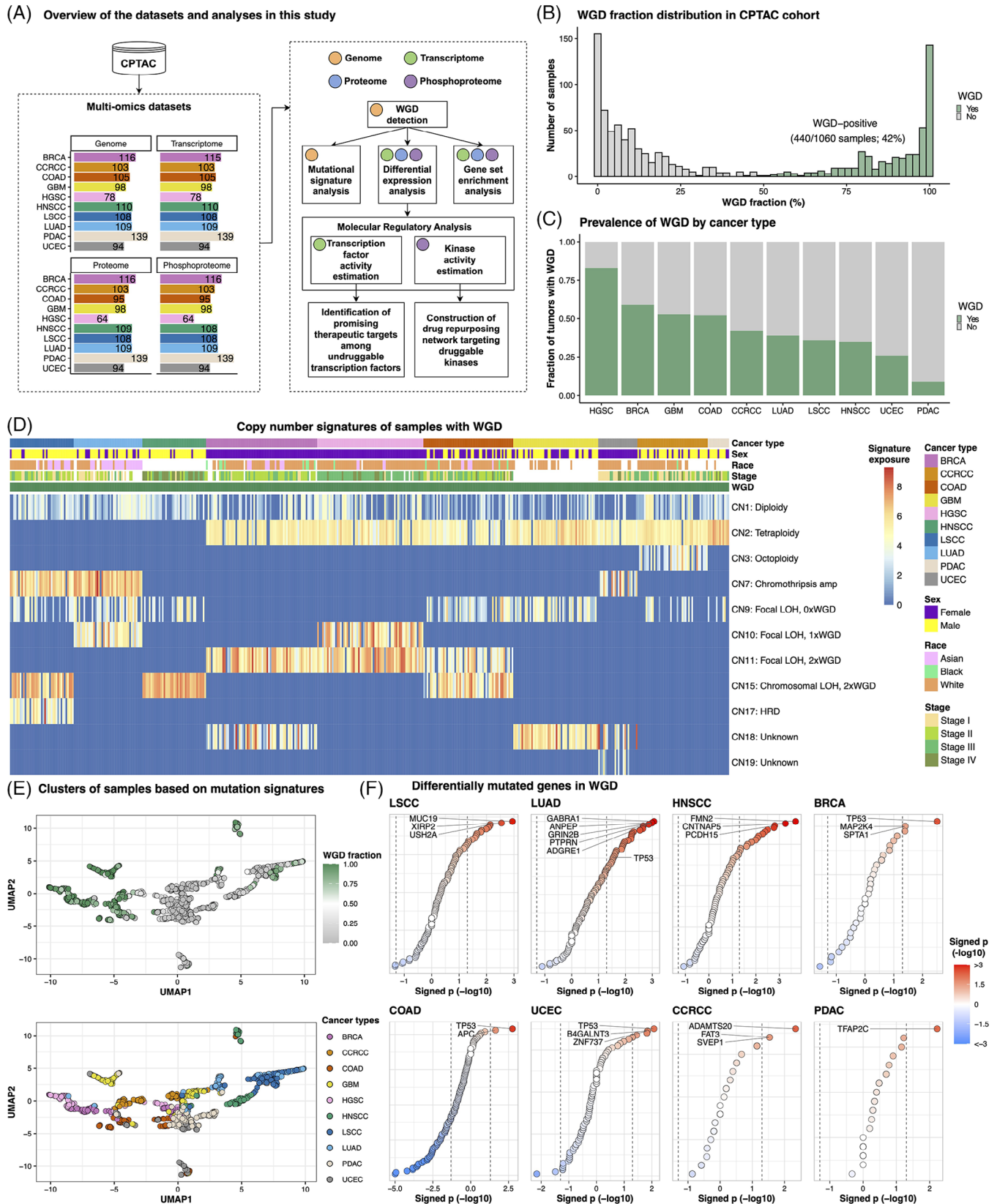


FIGURE 1 Copy number signatures associated with whole-genome doubling (WGD). (A) Schematic representation of the multi-omics datasets collected for each cancer type and the subsequent analysis pipeline. (B) Distribution of the WGD fraction within the Clinical Proteomic Tumour Analysis Consortium (CPTAC) 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) Uniform manifold approximation and projection (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. High-grade serous carcinoma (HGSC) and glioblastoma (GBM), which had no significantly enriched mutations in WGD, are not shown. The position of each dot along the x -axis and the colour of the dots indicate signed p -values obtained using Fisher's exact test. Only the top three significant genes were labelled in each cancer type, except for colon adenocarcinoma (COAD) and pancreatic ductal adenocarcinoma (PDAC), which had only two and one significant genes, respectively. In lung adenocarcinoma (LUAD), *PTPRN*, *GRIN2B*, *ANPEP* and *ADGRE1* genes exhibited the same p -values, and *TP53* was additionally labelled. BRCA, breast cancer; CCRCC, clear cell renal cell carcinoma; HNSCC, head and neck squamous cell carcinoma; LSCC, lung squamous cell carcinoma; UCEC, uterine corpus endometrial carcinoma.

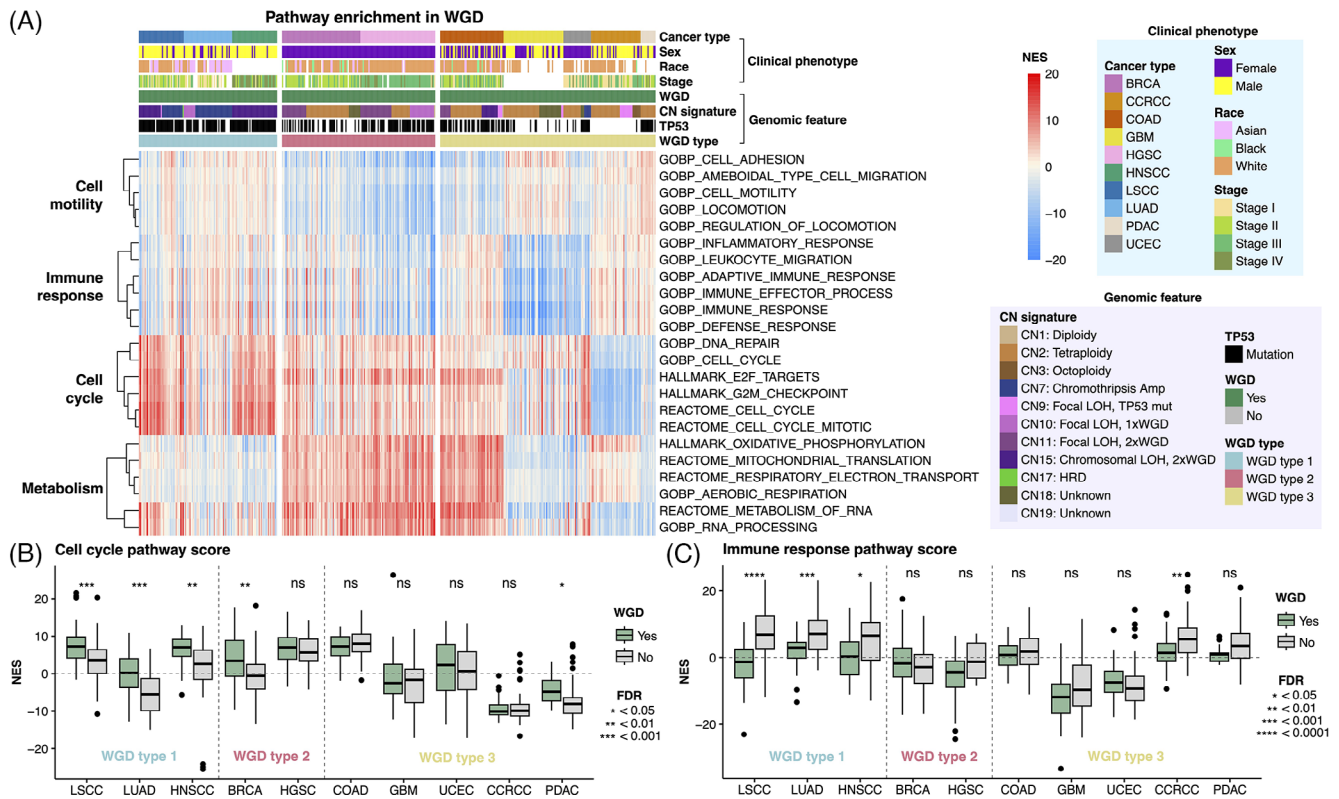


FIGURE 2 Pathway enrichment in whole-genome doubling (WGD)-positive tumours. (A) Normalised enrichment scores (NES) of pathways related to cell motility, immune response, cell cycle and metabolism in WGD-positive tumours. Pathways exhibiting significance (FDR < .05) and a log fold-change > .25 are shown. (B and C) Boxplot comparing NES scores of cell cycle (HALLMARK_E2F_TARGETS) and immune response pathways (GOBP_IMMUNE_RESPONSE) between WGD-positive and WGD-negative tumours in individual cancer types.

reported upregulation of the cell cycle and downregulation of immune response pathways in WGD,^{1,2,6} this pattern was specific to WGD type 1 (Figure 2B,C). WGD type 2 tumours were enriched for the dTTP metabolism pathway, which is responsible for DNA synthesis and maintenance, and the DNA endoreduplication pathway, which is a known mechanism inducing WGD⁷ (Figure S2A,B). WGD-positive COAD demonstrated significant activation of the Wnt signalling pathway, putatively attributed to *APC* mutation⁸ (Figure S2C).

To identify a potential therapeutic target for WGD tumours, TF activity was measured using the gene expression profile of WGD tumours. E2F family and MYC TFs were activated in pan-cancer WGD tumours, particularly in WGD type 1 (FDR < .1) (Figure 3A,B; Table S3A,B). Since many TFs have been deemed as ‘undruggable’,⁹ we

developed an integrative framework incorporating protein expression, cancer dependency and patient survival data to prioritise WGD-activated TFs as therapeutic targets (Figure 3C). Hence, BPTF was identified as a putative target for WGD-positive HNSCC tumours with poor prognosis (Figure 3D–F). Potential targets for WGD were found: *GLI2* (LSCC), *TFDP1* and *TFDP2* (LSCC, HNSCC, PDAC and HGSC), *E2F3* (LUAD) and *SREBF2* (BRCA) (Figure S3A–I; Table S3C). We further validated that the knock-down of *E2F3* in WGD-positive LUAD cell lines, and *BPTF*, *SFPQ* and *REST* in WGD-positive HNSCC cell lines resulted in a greater reduction in viability relative to WGD-negative cell lines (Figure 3G).

As kinases are well-known druggable proteins,⁹ we estimated kinase activity using proteomics and phosphoproteomics profiles of WGD tumours and constructed a

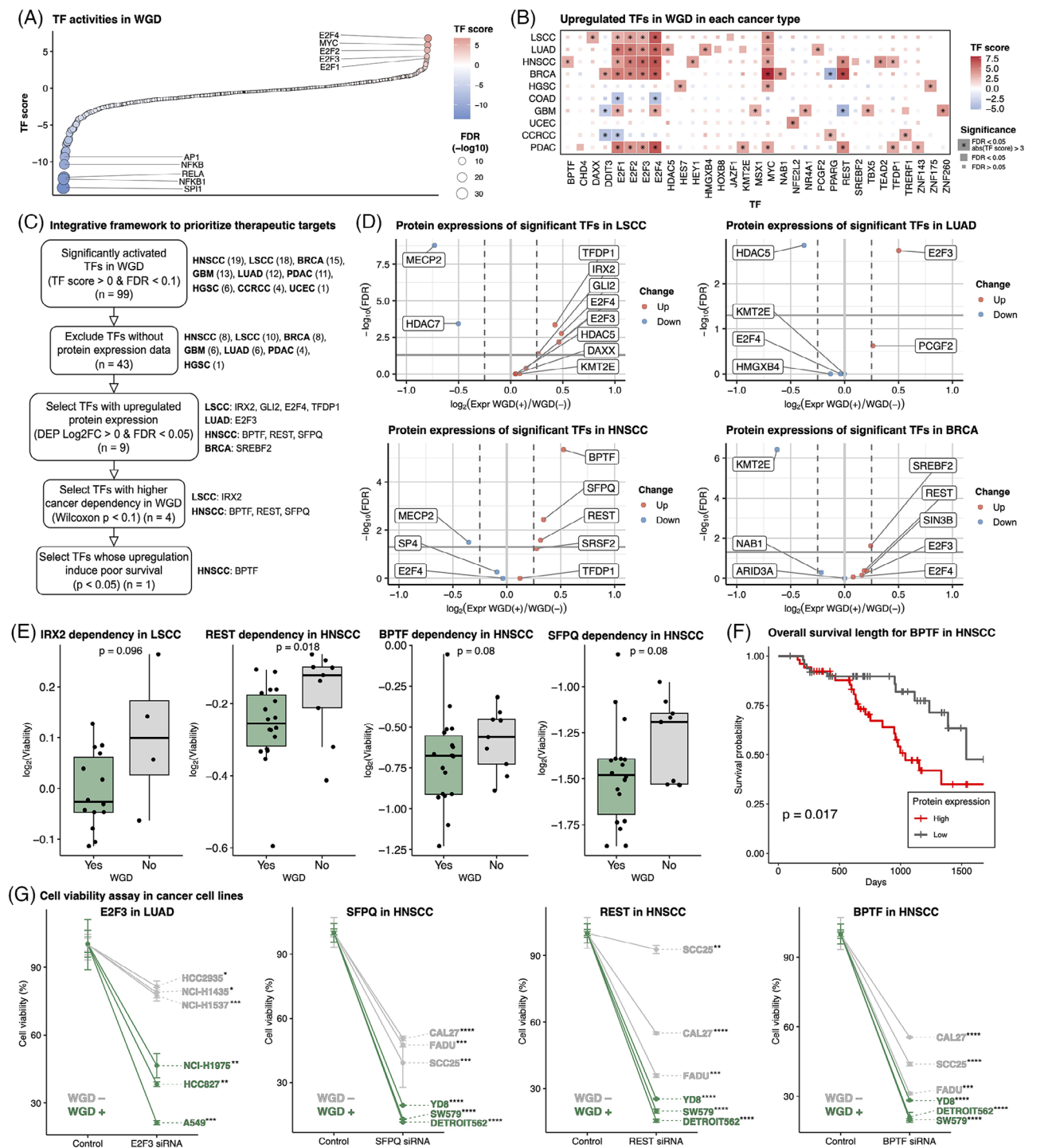


FIGURE 3 Determining potential therapeutic targets among activated transcription factors (TFs) in whole-genome doubling (WGD). (A) Dot plot depicting estimated TF scores in WGD-positive versus WGD-negative tumours across pan-cancer analysis. The top and bottom five TFs, ranked by TF score, are labelled. (B) Significantly upregulated TFs in WGD in each tumour type. The tiles are coloured based on the TF estimation score; TFs meeting the criteria of $FDR < .05$ 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 tumours versus WGD-negative tumours. (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 head and neck squamous cell carcinoma (HNSCC). (G) Cell viability assay for functional validation of identified TF targets. Specifically, E2F3 in lung adenocarcinoma (LUAD) and BPTF, REST and SFPQ in HNSCC were tested as these TFs were identified as promising targets and were experimentally available. Point indicates mean viability and error bar indicates standard deviation. p -Values from paired t -tests are denoted with asterisks. * $p < .05$, ** $p < .01$, *** $p < .001$ and **** $p < .0001$.

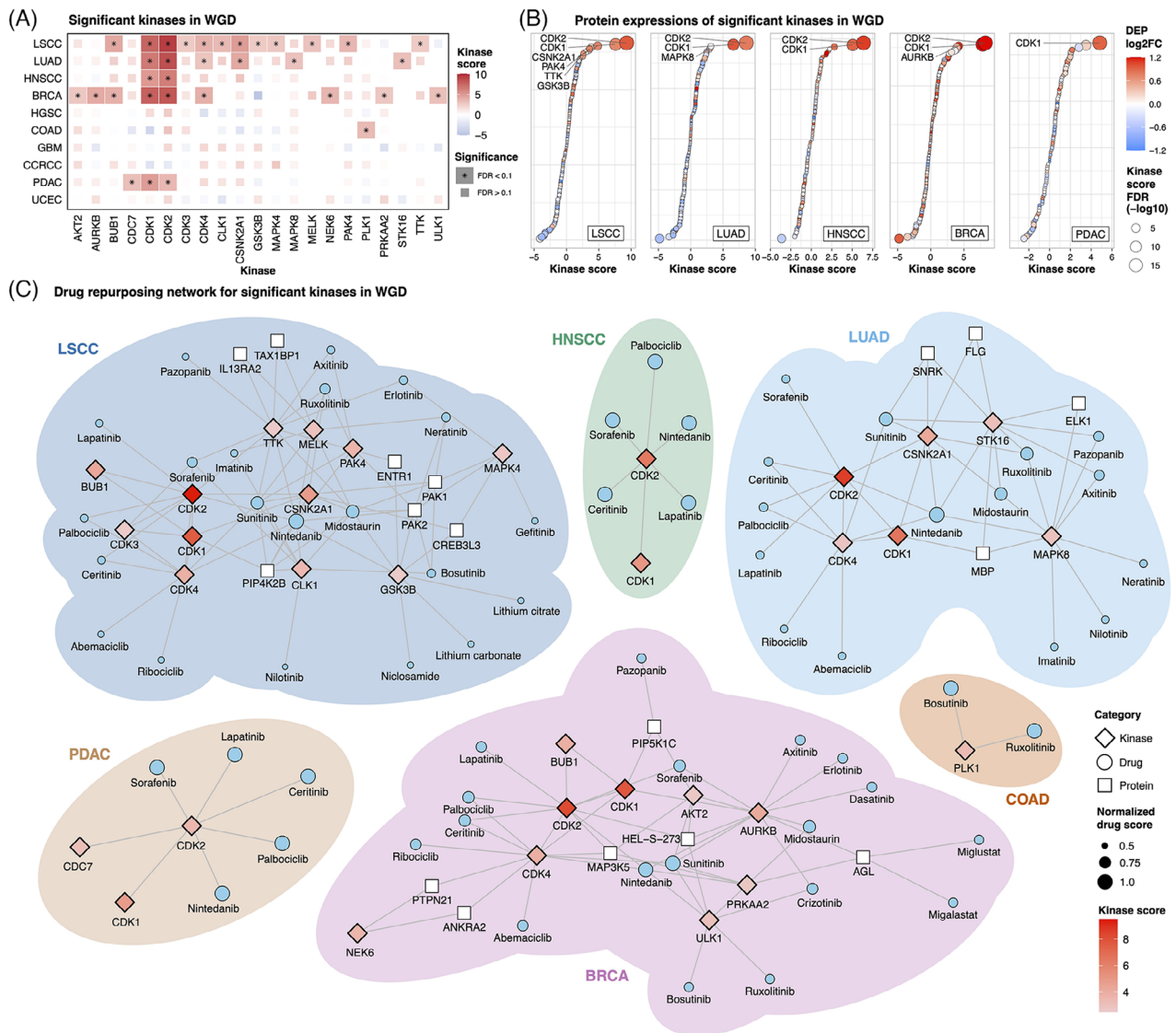


FIGURE 4 Drug repurposing strategies to target key kinases in whole-genome doubling (WGD). (A) Kinases with significant upregulation of expression in association with WGD in each tumour type. The tiles are coloured based on the kinase estimation score and kinases showing statistically significant upregulation are denoted with asterisks. (B) Dot plot depicting differentially expressed proteins among significantly upregulated kinases in WGD-positive tumours versus WGD-negative tumours. Six cancer types with significant kinase activation in WGD are displayed (lung squamous cell carcinoma [LSCC], lung adenocarcinoma [LUAD], head and neck squamous cell carcinoma [HNSCC], breast cancer [BRCA], pancreatic ductal adenocarcinoma [PDAC] and colon adenocarcinoma [COAD]), excluding COAD, which had no kinases with significantly upregulated protein expression. Kinases with significant upregulation of protein expression (FDR < .05 and log₂ fold-change > 0), as well as activation (FDR < .1), have been labelled. (C) Drug repurposing network for key kinases associated with WGD in each tumour type. Six cancer types with significant kinase activation of WGD are shown (LSCC, LUAD, HNSCC, BRCA, PDAC and COAD).

drug-repurposing network targeting kinases upregulated in WGD. We observed cancer-type-specific activations in WGD tumours, including CDK1/2 in cell cycle regulation, CSNK2A1 in DNA damage response, PAK4 in proliferation and apoptosis suppression, and PRKAA2/ULK1 in hypoxia-induced autophagy (Figure 4A,B; Table S4A). This indicates that WGD accelerates cell proliferation during tumourigenesis but also induces DNA damage and hypoxia, necessitating kinase activation to inhibit apop-

tos. The drug repurposing network analysis suggested potential drugs with significant selectivity in WGD, such as palbociclib (LSCC, BRCA), erlotinib, gefitinib and neratinib (LSCC) and nintedanib (PDAC) (Figures 4C and S4G-I; Table S4C).

In conclusion, our study presents the first proteogenomic analysis to characterise the molecular features of WGD across various cancer types. Three WGD types were defined: WGD type 1 (highly unstable genome), WGD type

2 (focal instability) and WGD type 3 (tetraploidy). Cell cycle pathway activation and immune response pathway inactivation were specific to WGD type 1. The integrative framework characterised a comprehensive molecular network from gene regulation to kinase–phosphorylation activity and suggested BPTF as a putative target for WGD-positive HNSCC. Further studies are warranted to evaluate the efficacy of drugs targeting WGD-specific TFs and kinases.

AUTHOR CONTRIBUTIONS

Eunhyong Chang and Joon-Yong An designed the study. Eunhyong Chang analysed the data. Su-Jung Kim conducted cell viability assay. Eunhyong Chang, Su-Jung Kim, Hee Sang Hwang, Kyu Jin Song, Kwoneel Kim, Min-Sik Kim, Se Jin Jang, Kwang Pyo Kim, Sungyong You and Joon-Yong An wrote and reviewed the manuscript. All the authors approved the final version of the manuscript.

[dataset] Li Y, Dou Y, Da Veiga Leprevost F, et al. CPTAC dataset. 2023. https://www.linkedomics.org/data_download/; <https://doi.org/10.1016/j.ccell.2023.06.009>

ACKNOWLEDGEMENTS

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT, the Korean government (NRF-2019M3E5D3073568 to Joon-Yong An, NRF-2019M3E5D3073567 to Kwang Pyo Kim, and 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 the 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 STATEMENT


Kwang Pyo Kim is the CEO of NioBiopharmaceuticals, Inc. Se Jin Jang is the Chief Technology Officer of SG Medical, Inc. All other authors report no conflicts of interest.

DATA AVAILABILITY STATEMENT

All CPTAC datasets we used are publicly available. 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.

ETHICS STATEMENT

Not applicable.

Eunhyong Chang^{1,2}
 Su-Jung Kim³
 Hee Sang Hwang⁴
 Kyu Jin Song^{3,5}
 Kwoneel Kim^{6,7} 
 Min-Sik Kim^{8,9,10}
 Se Jin Jang^{4,11}
 Sungyong You^{12,13}
 Kwang Pyo Kim^{3,5}
 Joon-Yong An^{1,2,14} 

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

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

³Department of Applied Chemistry, Institute of Natural Science, Kyung Hee University, Yongin, Republic of Korea

⁴Department of Pathology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea

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

⁶Department of Biology, Kyung Hee University, Seoul, Republic of Korea

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

⁸Department of New Biology, DGIST, Daegu, Republic of Korea

⁹New Biology Research Center, DGIST, Daegu, Republic of Korea

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

¹¹SG Medical, Inc., Seoul, Republic of Korea

¹²Department of Urology, Cedars-Sinai Medical Center, Los Angeles, California, USA

¹³Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, California, USA

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

Correspondence

Joon-Yong An, School of Biosystem and Biomedical Science, College of Health Science, Korea University, Seoul 02841, Republic of Korea.

Email: joonan30@korea.ac.kr

Kwang Pyo Kim, Department of Applied Chemistry, Institute of Natural Science, Kyung Hee University, Yongin 17104, Republic of Korea.

Email: kimkp@khu.ac.kr

ORCID

Kwoneel Kim  <https://orcid.org/0000-0002-4221-3421>

Joon-Yong An  <https://orcid.org/0000-0001-8839-6297>

REFERENCES

1. Bielski CM, Zehir A, Penson AV, et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet.* 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.* 2021;590(7846):492-497. doi:10.1038/s41586-020-03133-3
3. Li Y, Dou Y, Da Veiga Leprevost F, et al. Proteogenomic data and resources for pan-cancer analysis. *Cancer Cell.* 2023;41(8):1397-1406. doi:10.1016/j.ccell.2023.06.009
4. Steele CD, Abbasi A, Islam SMA, et al. Signatures of copy number alterations in human cancer. *Nature.* 2022;606(7916):984-991. doi:10.1038/s41586-022-04738-6
5. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* 2017;9(1):34. doi:10.1186/s13073-017-0424-2
6. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science.* 2017;355(6322):eaaf8399. doi:10.1126/science.aaf8399
7. Zeng J, Hills SA, Ozono E, Diffley JFX. Cyclin E-induced replicative stress drives p53-dependent whole-genome duplication. *Cell.* 2023;186(3):528-542. e14. doi:10.1016/j.cell.2022.12.036
8. 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.
9. Xie X, Yu T, Li X, et al. Recent advances in targeting the “undruggable” proteins: from drug discovery to clinical trials. *Signal Transduct Target Ther.* 2023;8(1):335. doi:10.1038/s41392-023-01589-z

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.