

LETTER TO THE JOURNAL

Deciphering protective genomic factors of tumor development in pediatric down syndrome via deep learning approach to whole genome and RNA sequencing

Childhood solid tumors represent a significant public health challenge worldwide, with approximately 15,000 new cases annually in the United States and an estimated 300,000 globally. Down syndrome (DS), a genetic disorder characterized by an extra full or partial copy of chromosome 21, results in distinctive developmental and physical features. Notably, individuals with DS exhibit a remarkable resilience against solid tumors compared to the general population, with an overall standardized incidence ratio (SIR) of 0.45, despite their increased susceptibility to hematologic malignancies [1]. This paradoxical observation has spurred extensive research aimed at uncovering the biological underpinnings of this natural resistance to solid cancers. Current theories suggest that the overexpression of specific genes on chromosome 21 may confer protective benefits (e.g. *RCAN1* contributes to antiangiogenic effects), and alterations in immune system function may enhance apoptosis and DNA repair pathways in individuals with trisomy 21 DS [2]. The well-established epigenetic effects of trisomy 21, which influence the entire genome, are another potential contributor to the reduced risk of solid tumors [3]. Nonetheless, these hypotheses face significant challenges, such as the potential oversimplification of complex genetic interactions and the lack of comprehensive genome-wide analyses. This study seeks to critically evaluate the correlations between genomic variants and cancer clinical phenotypes in patients with DS,

and proposes directions for future research into the genetic and molecular mechanisms that confer cancer resistance in DS, potentially transforming our understanding and treatment of pediatric cancers.

We conducted an innovative unbiased data-driven analysis in 2,452 whole-genome sequencing (WGS) samples with both DS individuals ($n = 635$) and pediatric oncology cases ($n = 280$) within the Gabriella Miller Kids First program project (<https://kidsfirstdrc.org/>) housed at the Children's Hospital of Philadelphia (Supplementary Figure S1). Additionally, 284 RNA sequencing samples from human peripheral blood mononuclear cells (PBMCs), a subset of WGS samples, were also analyzed, offering unprecedented insights into the complex interplay of genetic and immunological factors influencing cancer resistance.

1 | CANCER PROTECTIVE AND PREDISPOSING VARIANTS IN DS

The importance of each variant was calculated using deep learning algorithms, and their corresponding weights to DS cancer were generated based on linear algebra models as described in the Supplementary Materials and Methods. There were 2,523 unique cancer protective variants identified based on deep learning algorithms combined with linear algebra models in exonic, intronic, non-coding RNA and 5'untranslated region (5'UTR) regions. The prevalence for cancer protective variants in the DS cancer group (89.2%) is significantly higher compared to non-DS cancer individuals (58.1%) ($P = 1.11 \times 10^{-40}$), indicating that DS individuals may be protected against solid tumors by cancer protective variants identified in this study. The functional enrichment analysis revealed cancer development-related pathways for distinct categories of variants identified by WGS (Supplementary Figure S2). Of note, the functional terms differed between protective and predisposing variants, irrespective of the databases

List of Abbreviations: BD, birth defect; CAG, Center for Applied Genomics; DRAGEN, Dynamic Read Analysis for Genomics Bio-IT; DS, Down syndrome; EMR, electronic medical record; ICD, International Classification of Diseases; MLP, multilayer perceptron neural network; PBMC, human peripheral blood mononuclear cell; WGS, whole genome sequencing; VCF, variant call format; PD-L1, programmed death-ligand 1; PD-1, programmed cell death protein 1; EGFR, epidermal growth factor receptor; ID3, inhibitor of DNA binding 3; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PANTHER, Protein Analysis Through Evolutionary Relationships; WikiPath, wiki-based resource for collection, maintenance and distribution of biological pathways.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *Cancer Communications* published by John Wiley & Sons Australia, Ltd on behalf of Sun Yat-sen University Cancer Center.

used, including Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Protein Analysis Through Evolutionary Relationships (PANTHER), database of reactions, pathways and biological processes (Reactome), and wiki-based resource for collection, maintenance and distribution of biological pathways (WikiPath). A total of 121 genes exhibited overlapping between the cancer-protective and cancer-predisposing variants and their corresponding genes (Supplementary Table S1-S2). This intersection of genes, with a very low likelihood of occurring by chance ($P = 0.002$), attained statistical significance within cancer essential pathways such as the p53 pathway (False discovery rate [FDR] < 0.001) (Figure 1A). These outcomes aligned with expectations regarding the multifunctional roles of genes in the intricate processes of tumor development [4]. Variant type distributions revealed distinctive patterns within the 121 genes that are common to both categories (Figure 1B). Specifically, cancer-predisposing variants were more prevalent in exon regions for nonsynonymous or synonymous variants, in contrast, cancer-protective variants exhibited a higher prevalence in non-coding regions, suggesting regulatory roles in tumor development. Additionally, cancer-protective variants demonstrated earlier transcription activity than their predisposing counterparts (Figure 1C). These findings suggested that overlapping genes may serve dual roles as either cancer enhancers or suppressors, depending on the functional effects of the minor allele of a genetic variant. The nature of their impact is contingent upon various factors, including but not limited to variant categories and the genomic loci where these variants reside. To leverage prior cancer genetic knowledge, we integrated data from the Catalogue of Somatic Mutations in Cancer (COSMIC) and pediatric cancer driver genes to construct a curated list comprising 830 known cancer genes. Among the 121 overlapping genes, 18 were identified within this established cancer driver gene set. For instance, the isocitrate dehydrogenase 1 (*IDH1*) gene, associated with frequent mutations across various cancer types and tissues of origin [5], harbors a cancer-protective variant in the 5'UTR region (chr2:208254188-G-A) and a cancer-predisposing variant in the tail region (intron 8 of 9, chr2:208239246-A-G) (Figure 1D). Another example is gene inhibitor of DNA binding 3 (*ID3*), a member of the ID protein family implicated in cancer development, stemness, and metastasis [6], manifesting a protective variant in the 5'UTR (chr1:23559494-T-C) and a predisposing variant in the tail of exon 1 as a nonsynonymous mutation (chr1:23559171-C-G) (Figure 1D). Conversely, *UBQLN1*, a gene with cancer-protective variants located in the gene's tail (last intron) (Figure 1D), suppresses cancer stem cell-like traits in non-small cell lung cancer cells by regulating reactive oxygen species homeostasis [7]. For 148 cancer predisposing vari-

ants identified in cancer driver genes, the prevalence is 25% in DS cancer patients and 41.3% in non-DS cancer patients, suggested that DS population are under protections for cancer, and DS cancer patients may have different tumor development mechanisms compared to non-DS children.

2 | UP/DOWN-REGULATED GENES IN DS PATIENTS WITH CANCER

Analysis of the 284 RNA-sequencing (RNA-seq) PBMC samples yielded a highly valuable insight into the direction and magnitude of gene expression corresponding to selected variants. The categorization of cancer-protective and cancer-predisposing gene sets was further stratified into four subgroups for functional enrichment analysis based on direction. Notably, genes possessing cancer-protective variants that were down-regulated (suppressed) in cancer patients exhibited much stronger signals compared to other sets (Figure 1E-F). This phenomenon, especially evident in pathways relevant to tumor development such as Proteoglycans in cancer (FDR < 0.001) and Central carbon metabolism in cancer (FDR < 0.001), emphasized the significance of protective variants as dominant factors in tumor development among DS patients. A total of 1,785 genes with cancer-protective variants were presented, and genes with highest fold-changes in RNA-seq results while known as cancer driver genes were shown in Supplementary Table S3. Among 1,785 genes, 983 genes were found to be downregulated in cancer patients. Intriguingly, 86 of these down-regulated genes, mapped by 109 cancer-protective variants (Supplementary Table S4-S5), were identified as known cancer driver genes. These 86 suppressed genes in cancer patients with protective variants were not only associated with essential cancer pathways (FDR < 0.001) (Supplementary Figure S3A) but are also significantly enriched in cancer treatment response pathways, including epidermal growth factor receptor (*EGFR*) tyrosine kinase inhibitor resistance (FDR < 0.001) and programmed death-ligand 1 (PD-L1) expression and programmed cell death protein 1 (PD-1) checkpoint pathway in cancer (FDR < 0.001) (Supplementary Figure S3B-C). Referring to the National Cancer Institute's approved drug list, gene targets within the 86 genes are shown in Supplementary Table S6.

3 | ROLE OF CANCER-PROTECTIVE VARIANTS IN GENE REGULATION

Our study suggests that genes with cancer-protective variants down-regulated in cancer patients may act as a critical factor for the protective mechanism against solid tumors

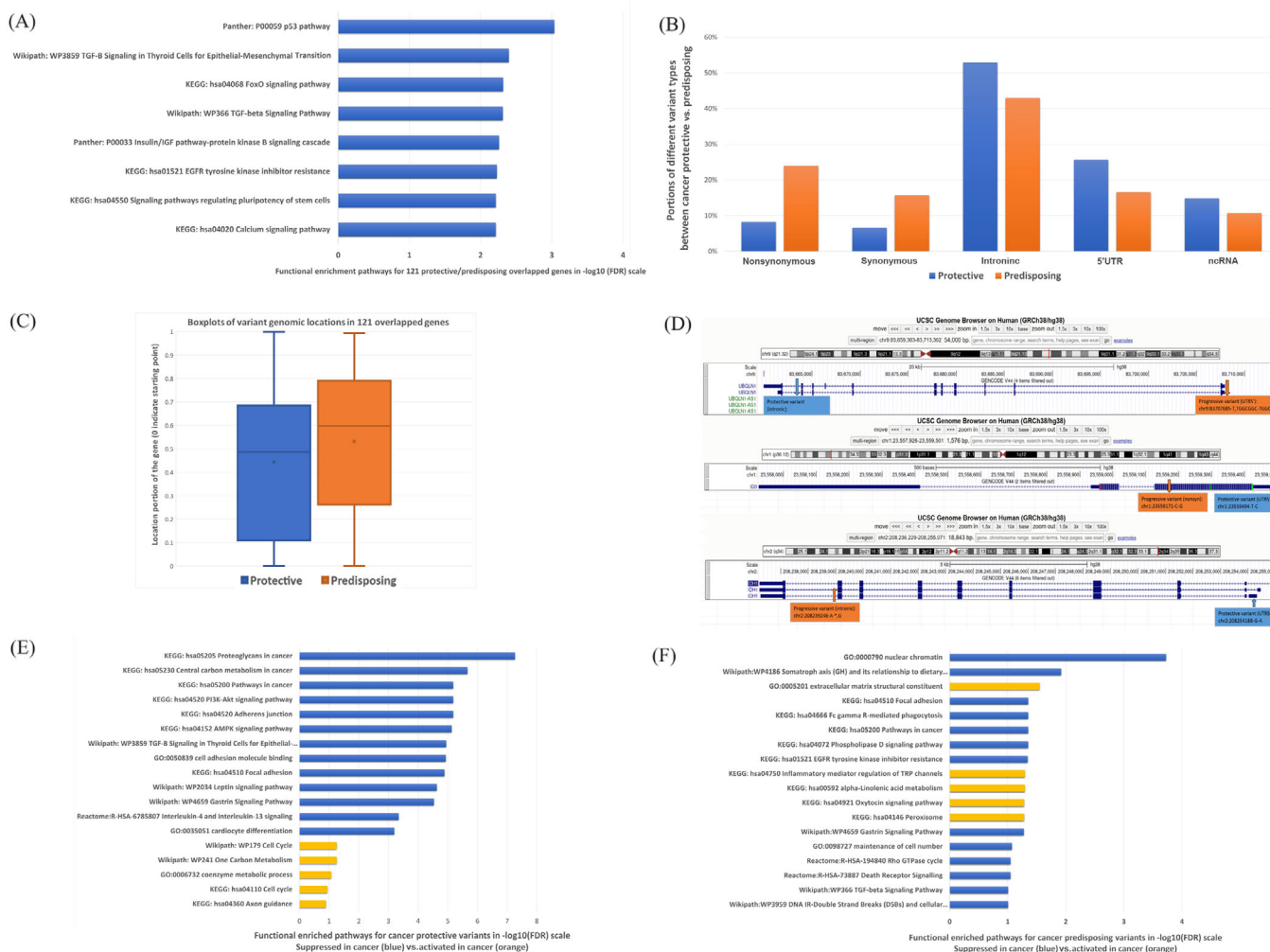


FIGURE 1 Functional analysis of cancer protective/ predisposing variants corresponding genes. (A) Functional enrichment for 121 corresponding genes with both cancer protective/predisposing variants: the enrichment analysis was performed using DAVID and WebGestalt, cancer essential pathways such as the p53 pathway were identified. (B) Proportions of different variant categories of cancer protective vs predisposing variants: bar graphs with the counts of variants in different genomic locations between cancer protective vs predisposing variants, variant type distributions revealed distinctive patterns for both categories. (C) Normalized locus of protective vs predisposing variants in overlapped genes: the locus was normalized based on transcript length and directions; cancer-protective variants demonstrated earlier transcription activity than their predisposing counterparts. (D) Examples of known cancer driver genes include *UBQLN1*, *ID3*, *IDH1*: these examples illustrate earlier genomic locus of cancer protective variants in known cancer driver genes. (E) Functional enrichment for genes corresponding to cancer protective variants: the enrichment analysis was performed using DAVID and WebGestalt, functional terms with blue color corresponding to genes suppressed in cancer patients based on RNA-seq DGE analysis and orange associated with genes active in cancer patients. (F) Functional enrichment for genes corresponding to cancer predisposing variants: the enrichment analysis was performed using DAVID and WebGestalt, functional terms with blue color corresponding to genes suppressed in cancer patients based on RNA-seq DEG analysis and orange associated with genes active in cancer patients. Abbreviations: DAVID, DAVID Bioinformatics platform; WebGestalt, WEB-based Gene Set Analysis Toolkit; *UBQLN1*, ubiquitin 1; *ID3*, inhibitor of DNA binding 3; *IDH1*, isocitrate dehydrogenase (NADP[+]) 1; DGE, differential gene expression.

in DS patients. A nonsynonymous cancer-protective variant (chr7:55205451-A-C) in *EGFR* exon 28, resulting in the truncation of the C-terminal domain of *EGFR*, has been previously reported to be associated with glioblastoma multiforme (GBM) patients [8]. This variant in *EGFR* demonstrated a high correlation coefficient (correlation coefficient > 0.7) with another nonsynonymous

cancer-protective variant (chr7:55205451-A-C) in exon 8 of *SEPTIN14*, and the *EGFR-SEPTIN14* fusion has been linked to glioblastoma with Icotinib-sensitive drug responses [9]. Furthermore, a variant (chr7:55255565-T-C) for the ncRNA *ELDR* (*EGFR* long non-coding downstream RNA) was identified in exon 1, with a previous study showing that knockdown of *ELDR* resulted in the downregulation

of *EGFR*, leading to the inactivation of downstream molecules, and it is considered a therapeutic potential target in cancer [10]. For the remaining genes down-regulated in cancer patients with cancer-protective variants that are not recognized as cancer driver genes, they also exhibit enrichment in cancer-related pathways, including the AMPK signaling pathway (FDR = 0.032), PI3K-Akt signaling pathway (FDR = 0.004), and Focal adhesion (FDR = 0.044) (Supplementary Figure S3D-F).

4 | UNLOCKING CANCER SOLUTIONS THROUGH DS GENETICS

This study significantly advances our understanding of how genetic factors associated with DS contribute to a reduced risk of solid tumor development. Through an examination of more than 2,000 WGS samples, we identified genetic variants playing important roles in either protecting or predisposing individuals to cancer. With the revealed correlations between protective variants, cancer mechanisms, and treatment response pathways, our findings warrant exploring new therapeutic interventions at the gene or pathway level. The development of targeted therapies, inspired by the natural protective mechanisms found in DS individuals, could transform the landscape of cancer treatment, with far-reaching implications extending beyond the DS population.

DECLARATIONS

AUTHOR CONTRIBUTIONS

Conceptualization and supervision: Yichuan Liu and Hakon Hakonarson. Literature search: Yichuan Liu. Data preparation and analysis: Yichuan Liu, Hui-Qi Qu, Xiao Chang, Frank D Mentch, Haijun Qiu, Kenny Nguyen, Kayleigh Ostberg, and Tiancheng Wang. Data interpretation: Yichuan Liu, Hui-Qi Qu, Xiao Chang, Joseph Glessner, and Hakon Hakonarson. Original draft writing: Yichuan Liu. Review and revision: Yichuan Liu, Hui-Qi Qu, and Hakon Hakonarson. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

Sample collection and biobanking for this study was supported by the Institutional Development Funds from the Children's Hospital of Philadelphia (CHOP) to the Center for Applied Genomics (CAG) and CHOP's Endowed Chair in Genomic Research (CAG). The sequencing data was provided through the Gabriella Miller Kids First Pediatric Research Program consortium (Kids First), supported by the Common Fund of the Office of the

Director of the National Institutes of Health (www.commonfund.nih.gov/KidsFirst), awarded to CAG. We express our gratitude to the CAG staff and the patients and families who generously contributed bio-samples to CAG/CHOP.

CONFLICT OF INTEREST STATEMENT

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING INFORMATION

The study was supported by the Institutional Development Funds from the Children's Hospital of Philadelphia to the Center for Applied Genomics, and The Children's Hospital of Philadelphia Endowed Chair in Genomic Research to Hakon Hakonarson.

DATA AVAILABILITY STATEMENT

The KidFirst data could be accessed at the Kids First Data Resource Portal (DRC, <https://portal.kidsfirstdrc.org/login>).

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

We confirm that all methods were carried out in accordance with relevant guidelines and regulations. All experimental protocols were approved by the Institutional Review Board (IRB) of the Children's Hospital of Philadelphia (CHOP) with the IRB number: IRB 16-013278.

CONSENT FOR PUBLICATION

Informed consent was obtained from all subjects. If subjects are under 18, consent was obtained from a parent and/or legal guardian with assent from the child if 7 years or older.

Yichuan Liu¹ 

Hui-Qi Qu¹ 

Xiao Chang¹

Frank D Mentch¹

Haijun Qiu¹

Kenny Nguyen¹

Kayleigh Ostberg¹

Tiancheng Wang¹

Joseph Glessner^{1,2,3}

Hakon Hakonarson^{1,2,3,4,5}

¹Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

²Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

³*Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA*

⁴*Division of Pulmonary Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA*

⁵*Faculty of Medicine, University of Iceland, Reykjavik, Iceland*

Correspondence

Yichuan Liu Ph.D, Center for Applied Genomics (CAG),
3615 Civic Center Blvd, Abramson Building, Philadelphia,
PA 19104, United States of America.

Email: liuy5@email.chop.edu

Hakon Hakonarson M.D., Ph.D., Center for Applied
Genomics (CAG), 3615 Civic Center Blvd, Abramson
Building, Philadelphia, PA 19104, United States of
America.

Email: hakonarson@email.chop.edu

ORCID

Yichuan Liu  <https://orcid.org/0000-0003-2023-072X>

Hui-Qi Qu  <https://orcid.org/0000-0001-9317-4488>

REFERENCES

1. Hasle H, Friedman JM, Olsen JH, Rasmussen SA. Low risk of solid tumors in persons with Down syndrome. *Genet Med*. 2016;18(11):1151-7.
2. Osuna-Marco MP, López-Barahona M, López-Ibor B, Tejera ÁM. Ten reasons why people with down syndrome are protected from the development of most solid tumors-a review. *Front Genet*. 2021;12:749480.

3. Satgé D, Bénard J, editors. Carcinogenesis in Down syndrome: What can be learned from trisomy 21? *Semin Cancer Biol*. 2008;18(5):365-71.
4. Pomerantz JH, Blau HM. Tumor suppressors: enhancers or suppressors of regeneration? *Development*. 2013;140(12):2502-12.
5. Pirozzi CJ, Yan H. The implications of IDH mutations for cancer development and therapy. *Nat Rev Clin Oncol*. 2021;18(10):645-61.
6. Lasorella A, Benezra R, Iavarone A. The ID proteins: master regulators of cancer stem cells and tumour aggressiveness. *Nat Rev Cancer*. 2014;14(2):77-91.
7. Jantrapirom S, Lo Piccolo L, Pruksakorn D, Potikanond S, Nimlamool W. Ubiquitin networking in cancers. *Cancers*. 2020;12(6):1586.
8. Cho J, Pastorino S, Zeng Q, Xu X, Johnson W, Vandenberg S, et al. Glioblastoma-derived epidermal growth factor receptor carboxyl-terminal deletion mutants are transforming and are sensitive to EGFR-directed therapies. *Cancer Res*. 2011;71(24):7587-96.
9. Zhu Y-c, Wang W-x, Li X-l, Xu C-w, Chen G, Zhuang W, et al. Identification of a novel icotinib-sensitive EGFR-SEPTIN14 fusion variant in lung adenocarcinoma by next-generation sequencing. *J Thorac Oncol*. 2019;14(8):e181-e3.
10. Sur S, Ray RB. Emerging role of lncRNA ELDR in development and cancer. *FEBS J*. 2022;289(11):3011-23.

SUPPORTING INFORMATION

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