

RESEARCH ARTICLE

Cancer carrier screening in the general population using whole-genome sequencing

Ya-Sian Chang^{1,2,3,4} | Dy-San Chao^{1,2,3} | Chin-Chun Chung¹ | Yu-Pao Chou^{1,2,3} | Chieh-Min Chang^{1,2,3} | Chia-Li Lin¹ | Hou-Wei Chu⁵ | Hon-Da Chen^{1,2,3} | Ting-Yuan Liu¹ | Yu-Hsuan Juan¹ | Shun-Jen Chang⁶ | Jan-Gowth Chang^{1,2,3,4,7} 

¹Center for Precision Medicine, China Medical University Hospital, Taichung, Taiwan

²Epigenome Research Center, China Medical University Hospital, Taichung, Taiwan

³Department of Laboratory Medicine, China Medical University Hospital, Taichung, Taiwan

⁴School of Medicine, China Medical University, Taichung, Taiwan

⁵Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan

⁶Department of Kinesiology, Health and Leisure Studies, National University of Kaohsiung, Kaohsiung, Taiwan

⁷Department of Bioinformatics and Medical Engineering, Asia University, Taichung, Taiwan

Correspondence

Jan-Gowth Chang, Center for Precision Medicine, China Medical University Hospital, 2 Yuh-Der Road, Taichung 404, Taiwan.

Email: d6781@mail.cmuh.org.tw

Funding information

China Medical University Hospital, Grant/Award Number: DMR-107-099

Abstract

Background: Cancer is a major cause of death, and its early identification and intervention have potential for clinical actionability and benefits for human health. The studies using whole-genome sequencing (WGS) and large samples analysis of cancer-related genes have been rarely done.

Methods: We performed WGS to explore germline mutations in coding and non-coding areas of cancer-related genes and non-coding driver genes and regulatory areas. Structural variants (SVs) was also analyzed. We used several tools and a subgrouping method to analyze the variants in 1491 healthy participants. Moreover, 275 cancer-related genes sequencing was carried out in 125 cancer patients.

Results: The incidence of familial cancer in the Taiwanese general population is 8.79% (131/1491). Cancer carrier rate of cancer-related genes is about 7.04% (105/1491) for pathogenic/likely pathogenic variants (P/LP) on ClinVar database only, and 28.24% (421/1491) for P/LP and loss of function variants. The carrier frequencies of cancer-related genes P/LP on ClinVar database were as follows: 8.40% (11/131), 7.11% (28/394), and 6.83% (66/966) in FC, 1MC, and nMC, respectively. The SVs and non-coding driver gene variants are uncommon. There are 1.54% (23/1491) of actionable cancer genes in American College of Medical Genetics and Genomics (ACMG), and the germline mutation rate of 275 cancer-related genes is 7.2% (9/125) in cancer patients including 4.0% (5/125) of actionable cancer genes in ACMG. After analyzing the frequencies of P/LP variants on *GJB2* and *SLC25A13* genes, we suggest that these two genes may not be cancer-related genes and need be re-evaluated.

Conclusions: WGS analysis can completely detect germline mutations in cancer carriers. This study use subgrouping approach for samples provides a strategy to study whether a gene or variant is a cancer-related gene or variant in the future studies.

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.

© 2022 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

KEYWORDS

GJB2, *SLC25A13*, variant analysis, whole-genome sequencing

1 | BACKGROUND

Cancer is a major public health problem worldwide and is the major leading cause of death in the world,^{1,2} and hereditary causes account about 3%–12.6% of adult cancers and 8.5%–10% of childhood cancers.^{3–6} The effect of hereditary factors on cancer development can be divided into low-penetrance susceptibility conferred by common germline variants (1.5–2.0-fold increase in relative risk), moderate penetrance (2.0–5.0-fold increase in relative risk), and high penetrance predisposition conferred by rare germline variants (>5.0-fold increase in relative risk).^{7–10} Moderate and high penetrance predispositions are usually have an autosomal-dominant inheritance pattern. The inherited genome can be interrogated at any stage of life, enabling prediction of the future risk of cancer.^{11–13} Screening for high-penetrance inherited variants is performed in families with clusters of tumor types, and more than 100 high-penetrance cancer predisposition genes (CPGs) are known.^{7,14–17} Many high-penetrance germline variants increase the risk of a broader range of cancers than classically described.^{6,7,11–17} High-penetrance pathogenic (P) variants are found in 5%–10% of unselected patients with cancer, but most inherited predispositions can be attributed to thousands of alleles common in the population that individually provide only a slightly increased risk of cancer.¹⁰ The best-characterized cancers now have more than 100 genomic regions associated with risk, accounting for more than 15%–20% of familial relative risk.¹⁰

Because cancer driver genes (CDGs) play a key role in cancer development,¹⁸ carriers of germline P/likely pathogenic (LP) variants of these genes will be at risk of cancer. This proposal is supported by the CPGs that are CDGs.^{19,20} CDGs variant-related familial cancer may be very rare and span many different genes beyond previous studies, and whole-genome sequencing (WGS) has not been applied to analyze variants of all CDGs, which may explain why germline P variants in many familial cancers have not been identified. More comprehensive methods are needed to detect carriers. In the cancer genome atlas (TCGA) cohort, 8% of adult cancer cases carrying P/LP germline variants in 152 CPGs.²⁰ The Pan-Cancer Analysis of whole Genomes study of the TCGA and International Cancer Genome Consortium identified that 17% of all patients had rare germline variants associated with cancer.²¹ No matter TCGA or ICGC, they were derived from cancer tissues not from the general population. Recently, Rheinbay

et al. analyzed driver point mutations and structural variants (SVs) in non-coding regions across 2658 genomes of different cancers, and their results showed that point mutations and SVs were less frequent in non-coding genes and regulatory sequences than in protein-coding genes.²² The roles of germline variants of these non-coding genes and regions in the development of familial cancer or cancer need to be confirmed.

Variants of cancer-related genes can be detected by whole-exome sequencing,^{20,23} but this will miss genes with only non-coding driver variants in cancer-related non-coding genes or SVs. WGS can not only identify these non-coding regions for potential driver events but can also explore changes in non-coding driver genes.²² WGS can also localize SVs, breakpoints, and connections between distinct genomic loci (juxtapositions).

In this study, we collected 152 CPGs,²⁰ 299 CDGs,²³ 568 CDGs,²⁴ and the non-coding driver genes or regions of Rheinbay's study,²² totally 724 protein coding genes and 36 non-coding driver genes or regulatory areas were included. We also analyzed the germline variants of white blood cells (WBCs) in 125 cancer patients using a 275 cancer genes-panel to explore the germline variant frequency in cancer patients.

2 | METHODS

2.1 | Study participants

Thousand four hundred and ninety one participants' data were collected from the Taiwan Biobank (TWB), a general population-based research database comprising cancer-free residents aged 30–70years enrolled at 31 recruitment stations in Taiwan since 2008. Details on the TWB can be found on its official website (<https://taiwanview.twbiobank.org.tw/index>).^{25–27} We subgroup the participants according to the number of cancer cases in the family as FC: more than 2 of 1st degree family members with cancers (131 cases); 1MC: one family member with cancer; nMC (394 cases): no family member with cancer (966 cases). The study was conducted with the approval of the Institutional Review Board (IRB) of China Medical University Hospital (CMUH 108-REC1-091), of the IRB on Biomedical Science Research/IRB-BM Academia Sinica, Taiwan (TWBR10809-02). All participants were self-reported as healthy and Han Chinese ancestry, and signed the written informed consent forms.

We retrospectively reviewed the sequence data from 125 cancer patients who underwent genetic testing from January 2017 to July 2021. This study was approved by the IRB of the China Medical University Hospital (CMUH106-REC1-047).

2.2 | DNA extraction

Peripheral blood of enrolled participants was collected into sodium citrate tubes and DNA was isolated using a Chemagic™ Prime™ instrument. DNA length was measured using a Fragment Analyzer (Agilent) and purity was assessed by measuring the optical density (OD) at 260/280 nm. Samples with an OD 260/280 ratio of 1.6–2.0 were considered pure.

2.3 | A cancer panel including 275 cancer-related genes for detecting germline mutation in WBCs

We used a commercialized kit (DHS-3501Z) from QIAGEN Co. (QIAseq Targeted DNA Panel, Human Comprehensive Cancer Panel), which includes 275 cancer-related genes to analyze the DNAs of WBCs using next-generation sequencing (NGS) for germline mutations of cases with solid cancers. Totally, there are 125 cases of solid cancers including 40 lung, 30 ovarian, 16 colon, 14 breast, 4 stomach, 3 each with endometrial and urothelial cancers, and other cancer types with less 3. The clinicopathological features of the patients are given in [Table S1](#).

2.4 | Bioinformatics analysis of variants

We used WGS data released by the Taiwan Biobank (<https://taiwanview.twbiobank.org.tw/search>).^{25,26} The VCF data were analyzed in our bioinformatics pipeline, which uses the HGVS database to localize the variant in RNA, and protein and official name: (<https://annovar.openbioinformatics.org/en/latest/user-guide/download/>) and (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/>). We used dbSNP to get an official number (<https://ftp.ncbi.nlm.nih.gov/snp/archive/b153/VCF/>). We used the genomAD (<https://gnomad.broadinstitute.org/>) and Taiwan biobank (https://www.twbiobank.org.tw/new_web/) databases, and our database, to evaluate the frequencies of variants, and we analyzed variant character using the ClinVar database (https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh38/),²⁸ and the analytic tools CADD (https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh38/),^{29,30} RegulomeDB ([\[spotter.org/\]\(http://spotter.org/\)\),³¹ and FunSeq2 \(<http://funseq2.gersteinlab.org/>\).³² We leveraged Decipher \(<https://www.deciphergenomics.org/about/downloads/data>\) and OMIM \(<https://www.omim.org/downloads>\) \(Licensing and Registration\) to explore the relationships of the variants detected with disease. We also used American College of Medical Genetics and Genomics \(ACMG\) guideline to analyze the reporting genes.³³](https://www.regulation</p></div><div data-bbox=)

2.5 | Statistic testing of rare variants of cancer-related genes for nMC, 1MC, FC, and 688 cancerous tissues

The gene or variant-based statistic testing of cancer-related genes in four subgroups were performed for the aberrant frequencies of deleterious variants in nMC, 1MC, FC, and cancer tissues using method as following: We analyzed the variants in the coding regions in cancerous cases and controls considering the 2×2 contingency and those variants whose CADD score >25 were further analyzed to study the association of variants in cancer. We calculated the relative risk of variant in cases as compared to control groups to evaluate the level of enrichment in cancer. The risk ratio was defined to be $RR = \frac{a/(a+b)}{c/(c+d)}$ and adjusted chi-square test was used to identify significant associations. The R-package of “epitools” was used to calculate RR of each variant in our study.

2.6 | Confirmation of variants by PCR and direct sequencing

Primers for direct sequencing validation were designed using Primer3 software. The PCR primers used are shown in [Table S2](#). PCR amplifications were performed using ProTag Plus DNA Polymerase (Protech Technology Enterprise) following the manufacturer's instructions and our previous study.³⁴

3 | RESULTS

The strategy, participants, targeted genes, and analyzing approaches of this study are shown in the [Figure 1](#). Totally, 1491 healthy participants and 125 cancer patients were included, and the healthy participants were further subclassified into FC, 1MC, and nMC according to the cancer occurrence in the family member, and then they were subjected for 724 cancer-related genes, and 36 non-coding driver genes and regulatory areas ([Figure 1A](#)). The relationships among 152 CPGs, 299 CDGs, and 568 CDGs are shown in [Figure 1B](#). The detailed relationships

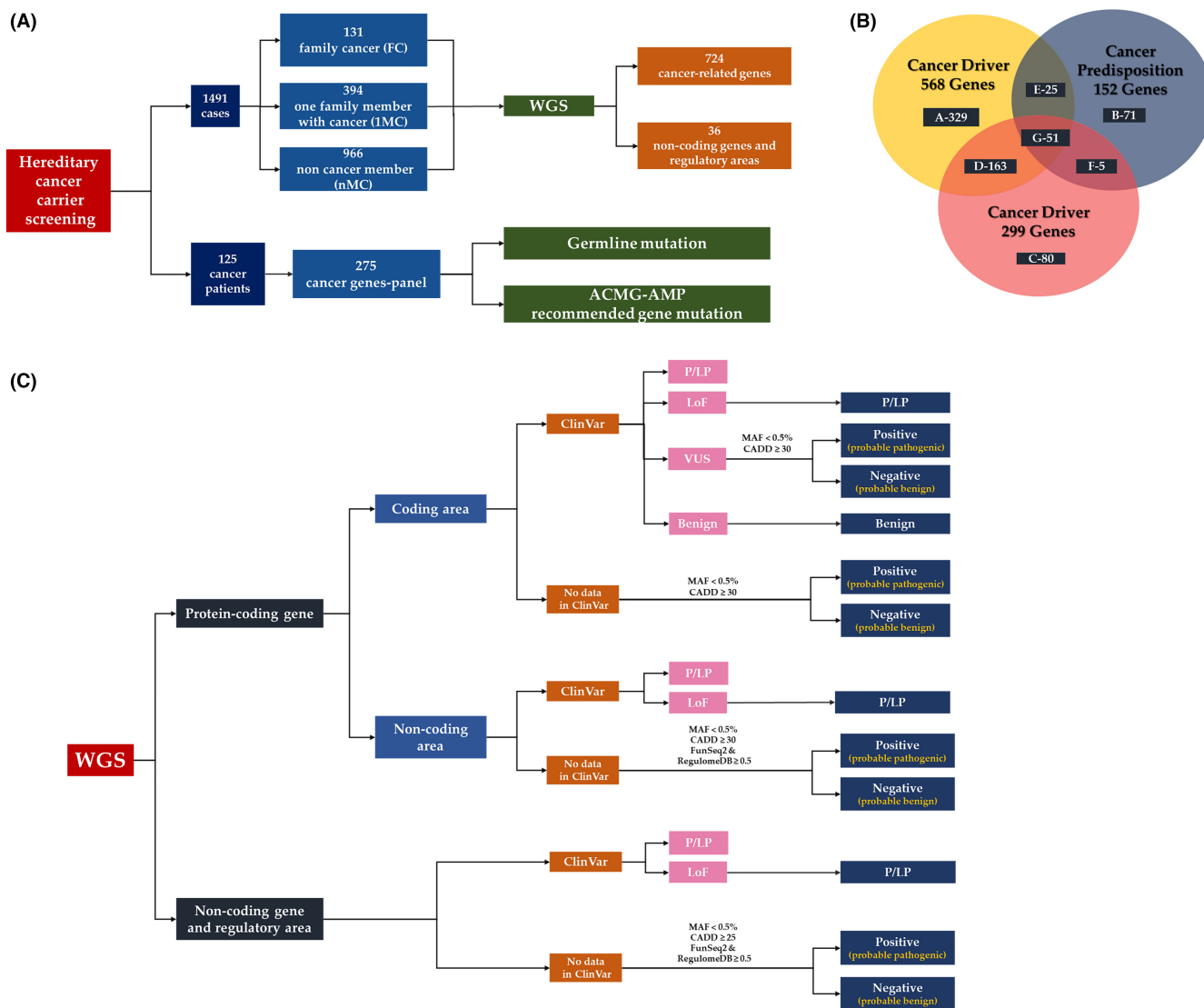


FIGURE 1 The strategies and analyzing methods of this study are shown. (A) The strategy, participants, targeted genes, and analyzing approaches are shown. (B) Relationships among 152 cancer predisposition genes, 299 cancer driver genes, and 568 cancer driver genes are shown. (C) WGS analysis methods. We analyzed variants of coding and non-coding areas in protein-coding genes, and non-coding genes and regulatory areas using the ClinVar database, and prediction tools (e.g., CADD, RegulomeDB, and FunSeq2) to analyze null variants. WGS, whole-genome sequencing.

of these panel genes are shown in [Table S3](#). The strategies of WGS analysis are shown in [Figure 1C](#).

We analyzed variants of coding areas and non-coding areas including introns, promoters, and regulatory elements using the ClinVar, TCGA, and COSMIC databases for known variants, and we used three prediction tools, CADD for coding areas, and CADD, RegulomeDB, and FunSeq2 for non-coding areas, to analyze null variants on ClinVar database. For SVs, we used ClinVar and Decipher databases to explore the clinical significance.

We select the targeted variants using ClinVar P/LP, or CADD ≥ 30 and MAF < 0.5% for coding areas, and CADD ≥ 30 and MAF < 0.5% and FunSeq2 ≥ 0.5 and RegulomeDB ≥ 0.5 for non-coding areas as a selected criteria.

3.1 | Stratification of participants

We collected 1491 WGS data from the general population of Taiwan Biobank. The male to female ratio was near 1 (744/747), and the age range was 30–70 years, and the detail of demographic data was shown in [Table S4](#).

3.2 | Variant detection of coding areas and non-coding areas in 724 cancer-related genes

Totally, 724 cancer-related genes (152 CPGs + 299 CDGs + 568 CDGs and excluding the overlap genes) were

collected in this study, and the summarized results which include P/LP variants in the ClinVar database, or CADD score ≥ 30 are shown in the [Table 1A](#). The detailed variants are shown in [Tables S5–S8](#).

In the coding areas of 724 cancer-related genes, 74 participants have ClinVar P/LP ([Table S5](#)), and 573 participants have CADD score ≥ 30 variants with no ClinVar data which we consider as possible pathogenic (PP) variants ([Table S6](#)). Among the 573 participants, 177 have loss of function variants.

In the non-coding areas of 724 cancer-related genes, there are 45 participants who have ClinVar P/LP ([Table S7](#)), and 300 participants have CADD score ≥ 30 variants with no ClinVar data which we consider as PP variants ([Table S8](#)). Among the 300 participants, 139 participants have CADD ≥ 30 variants resulting from disrupting splicing machinery.

From these results, we suggest the cancer carrier frequency of driver genes of Taiwanese general population is 7.04% (105/1491) for ClinVar P/LP, or 28.24% (421/1491) for all the important variants including P/LP, stop gain or loss, frameshift, and splicing-disrupted ([Table 1A](#)). In these frequencies, we do not include the SVs and non-coding genes driver variants that may result in slightly underestimating the frequencies.

The variants of P/LP of ClinVar in 724 cancer-related genes in FC, 1MC, and nMC are shown in [Table 1B](#). The frequencies of several P/LP variants of *GJB2* gene were higher in 1MC and nMC than FC (p.Leu79X variant, 0.76%, 1.27%, and 1.35% for FC, 1MC, and nMC, respectively; for c.761G>GCGTT and c.549CAT>C frameshift variants were found only in 1MC and nMC) and the frequency of the P variant Val37Ile was very high among FC, 1MC, and nMC ($>16\%$ in all groups) ([Table S5](#)). For *SLC25A13*, the frequency of c.1043TCATA>T was 1.53%, 1.52%, and 1.55% for FC, 1MC, and nMC, respectively, and there was no difference in frequency among FC, 1MC, and nMC. We used statistic test for these two genes, and the results showed there are no significant different between FC, nMC or 1MC. Similar results were found, after we further used 688 different cancer tissues to confirm the results ([Table 2](#); [Table S9](#)). Therefore, these two genes are not likely cancer-related genes.

A ClinVar P variant in the *SBDS* gene had a $>0.2\%$ allelic frequency in gnomAD and our three groups. Therefore, this variant is not a moderate- or high-risk variant, and not included in the analysis of cancer-related gene carrier frequency ([Table S7](#)). After removing these variants, the cancer-related gene carrier frequencies of P/LP on ClinVar database were 8.40% (11/131), 7.11% (28/394), and 6.83% (66/966) in FC, 1MC, and nMC, respectively ([Table 1B](#)).

Several ClinVar P/LP variants were found in the non-familial cancer groups, such as *ATM* c.1787CAA>C,

p.K468X, and *RAD51D*, c.531T>TTA, p.K111IX, in 1MC and nMC; *BRCA2* c.8242C>T, p.Ser2670Leu, and *MSH6* c.3378C>T, p.Arg1076Cys, in 1MC; *DNMT3A* c.2982G>A, p.Arg882Cys, *BLM* c.1653G>GA, p.G512GX, *FANCA* c.1018TGTGA>T, p.TH329X, *FANCI* c.1926C>T, p.Arg614*, *RNF213* c.14572G>A, p.Arg4810Lys, and *WRN* c.748TA-A>T, p.K167X, in the nMC ([Table S5](#)). Therefore, we suggest that cancer carrier screening should not be limited to FC, and must include other non-familial cancer groups. Additional, several P/LP variants were found only in the FC group, such as *JAK2* c.2343G>T, p.Val617Phe, and *LZTR1* c.379C>CG, p.T7TX ([Table S5](#)), which may be causal variants.

We also evaluated the carrier frequency of reporting of secondary findings in ACMG, and the results showed that the frequency of P/LP of ClinVar in 1491 participants is 1.54% (23/1491) for 28 reporting cancer genes in ACMG-AMP guideline ([Table S10](#)).³³

3.3 | Analysis of non-coding genes and regulatory areas

In the non-coding driver genes and regulatory areas of Rheinbay's study, there are 13 participants having variants CADD score ≥ 25 ([Table S11](#)). The results were shown that 13 variants were found including three at *Hes1-1*, five at *HIST1H2AM*, and one at *SDCCAG8-3* promoters, respectively, and four at lncRNA *RMRP*. The promoter variant of *SDCCAG8-3* has a high CADD score, but has conflicting interpretation on ClinVar database. These variants may be not cancer-related.

3.4 | SV analysis in cancer-related genes

We found 2002 SV events in the 724 cancer-related genes from 1488 NGS samples and selected rare SVs (those in ≤ 2 samples), that involve 603 oncogenes (209 deletions, 43 duplications, 339 insertions, and 12 inversions) and 876 tumor suppressor genes (284 deletions, 66 duplications, 508 insertions, and 18 inversions) ([Table S12](#)). Summary and details of rare SVs in subgroups are shown on [Table S13](#), and most of them are located on the intronic regions. We specifically focused on the six SVs that involved exon areas of targeted genes, including three oncogenes, two tumor suppressor genes, and one biphasic gene ([Table S14](#)). These function disrupted exon-related SVs may play a role as cancer driver, which need to be further studied.

3.5 | Results of a special cancer family

We asked a participant with six family members with cancer spanning three generations for consent

TABLE 2 The statistics of the variants with ClinVar P/LP or CADD ≥ 25 between cancer and other non-cancer groups

Gene name	dbSNP	Cancer versus nMC (p-value)	Cancer versus nMC (adjusted p-value)	Cancer versus IMC (p-value)	Cancer versus IMC (adjusted p-value)	Cancer versus FC (p-value)	Cancer versus FC (adjusted p-value)
<i>GJB2</i>	rs776335807	NA	NA	0.1953612	0.716	NA	NA
<i>GJB2</i>	rs111033204	0.7168588	0.765	0.8846106	0.885	0.5281364	0.658
<i>GJB2</i>	rs80338943	0.8147019	0.815	0.814702	0.885	0.65822	0.658
<i>GJB2</i>	rs587783644	0.4083326	0.503	NA	NA	NA	NA
<i>SLC25A13</i>	rs80338725	0.006740853	0.108	0.08331204	0.716	0.3176162	0.658
<i>SLC25A13</i>	rs752235032	0.4083326	0.503	NA	NA	NA	NA
<i>SLC25A13</i>	rs139149160	0.08690023	0.503	0.2740806	0.754	0.5281364	0.658
<i>SLC25A13</i>	rs80338722	0.4083326	0.503	0.1953612	0.716	NA	NA
<i>SLC25A13</i>	rs776461118	0.4083326	0.503	NA	NA	NA	NA
<i>SLC25A13</i>	rs949468946	0.2262096	0.503	0.4395195	0.802	0.655743	0.658
<i>SLC25A13</i>	rs398122839	0.2262096	0.503	0.7119804	0.885	0.2027535	0.658
<i>SLC25A13</i>	rs1484296612	0.4083326	0.503	NA	NA	NA	NA
<i>SLC25A13</i>	rs80338720	0.3988902	0.503	0.5106532	0.802	0.6458937	0.658
<i>SLC25A13</i>	rs751343245	0.2421313	0.503	NA	NA	0.02470359	0.247
<i>SLC25A13</i>	rs879255503	0.4083326	0.503	NA	NA	NA	NA
<i>SLC25A13</i>	rs80338717	0.6843908	0.765	0.7706653	0.885	0.2053228	0.658
<i>SLC25A13</i>	rs1254503252	0.2262096	0.503	0.4395195	0.802	0.655743	0.658

Abbreviation: P/LP, pathogenic/likely pathogenic.

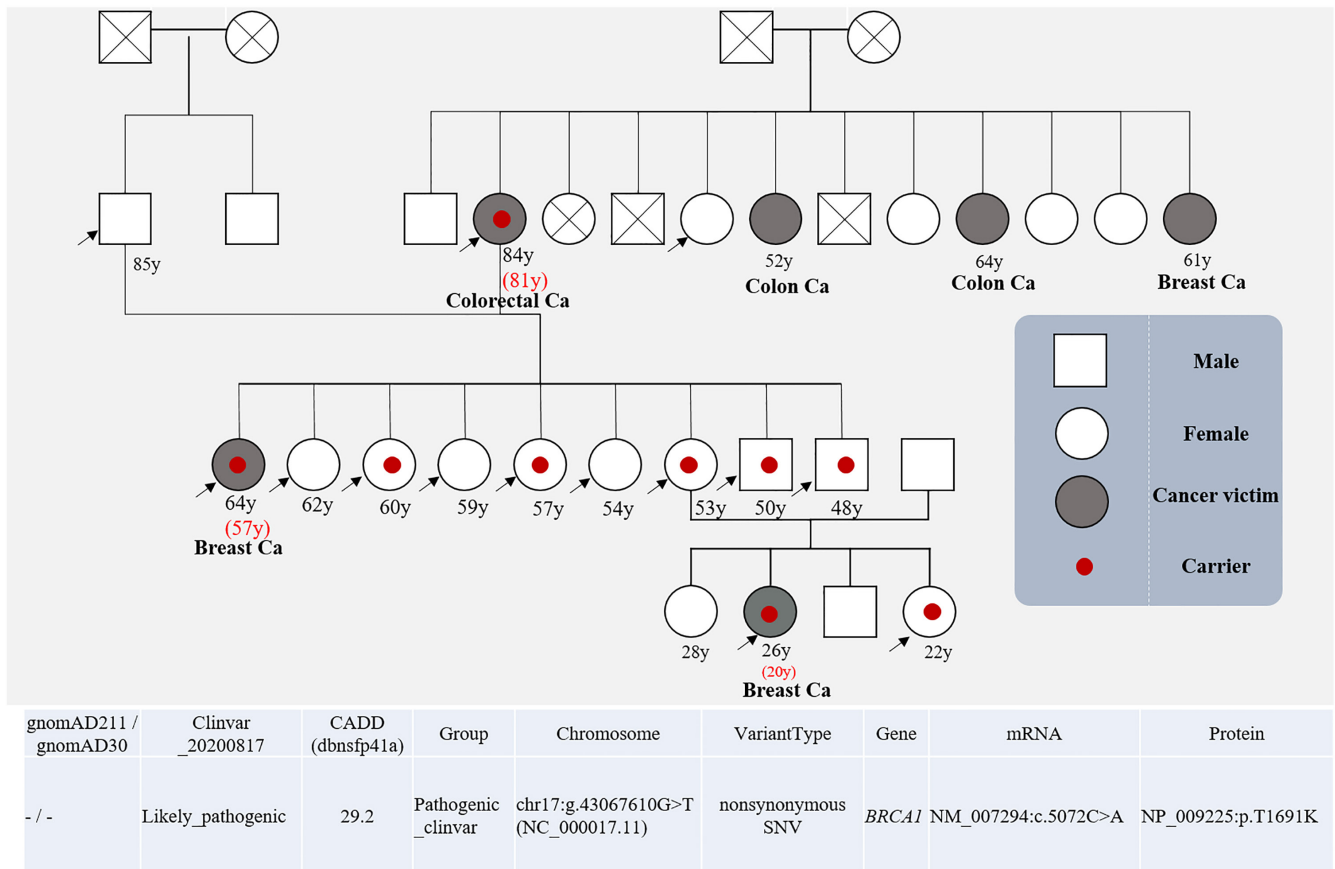


FIGURE 2 Pedigree and genetic analysis results of a cancer family. The arrow indicates the members who took WGS analysis. Red numbers are the times of cancer detection. Red dot is carrier cases under WGS analysis. WGS, whole-genome sequencing.

to analyze his family. We performed WGS of 14 members of this family and found nine with the c.5072C>A, p.T1691K mutation on the *BRCA1* gene (Figure 2). This family had two breast cancer members including one at a young age (20 years old), with triple negative breast cancer, one with colorectal cancer, and three females >50 years old with no cancer. From these results, we suggest that other factors may play a modification role in the time of cancer development.

3.6 | Evaluation of germline mutation of cancer patients

We explored the germline mutations of patients with solid cancer using a cancer panel with 275 cancer-related genes, and this panel contains 22/28 of reporting of secondary findings in ACMG-AMP.³³ The results showed that 7.2% (9 of 125) cancer patients have mutations on 275 cancer-related genes, and 4.0% (5 of 125) have reporting of secondary findings in ACMG-AMP (Table S15).

3.7 | Confirmation of base and SVs by PCR and direct sequencing

We sequenced >20% of the variants to confirm the NGS results (Figure 3), and we also confirmed the SVs using Sanger sequencing (Figure 4). The results showed that the false positive rates were about 5% for single base variants and 25% for SVs. From these results, we suggest that the confirmation of driver variants is needed.

4 | DISCUSSION

CDGs is the major determinant of cancer, and the number differs according to the approach used. In total, we analyzed 724 cancer-related genes, and 36 non-coding driver genes and regulatory areas including the studies of Rahman et al.,¹⁹ Bailey et al.,²³ Rheinbay et al.,²² and Martinez-Jimenez et al.²⁴ to detect carriers of cancer-related gene mutations using GS, and our results showed that this approach is a more comprehensive panel and may not lose reporting cancer genes in ACMG, and our approach is a WGS-based method which can cover all the

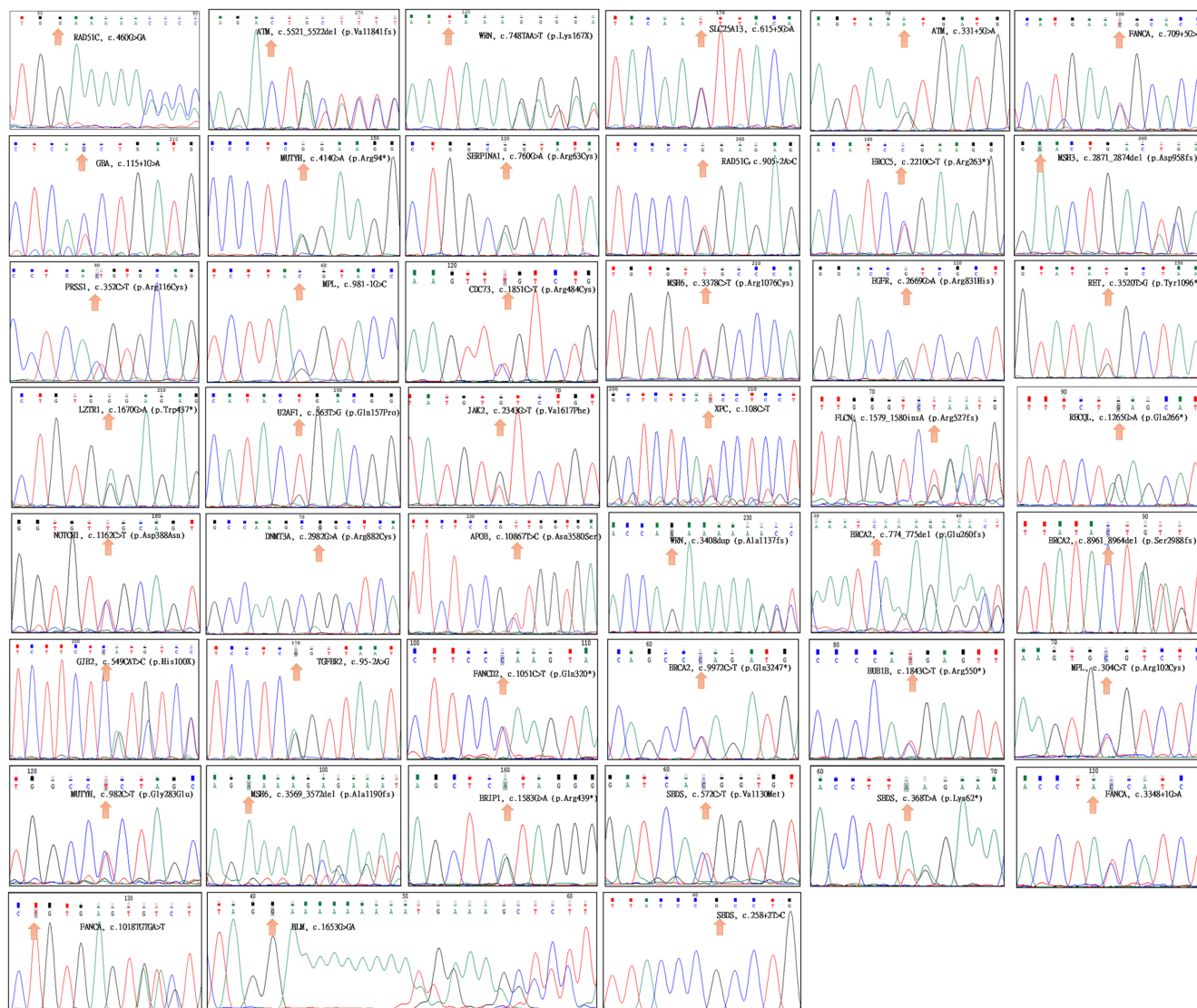


FIGURE 3 Sanger sequencing of pathogenic or likely pathogenic variants.

cancer-related genes for the new changes of ACMG-AMP in the future.³³

The frequencies of several detrimental variants of cancer-related genes (such as *GJB2* and *SLC25A13*) were higher in the non-familial cancer group than the familial cancer group. We used the data of subgrouping, and found no differences between FC and non-FC, and suggest these genes or variants may not play a role in the cancer development or not a true cancer-related genes. We further used statistic test and cancer tissues to confirm the subgrouping findings. Therefore, we suggest that *GJB2* and *SLC25A13* may not be cancer-related genes according to the results of statistic test. We suggest this subgrouping approach could be used to screening whether a gene or variant is a cancer-related gene or variant, and then using statistic test to confirm the finding. The power of screening will be increased after collected more data in the subgroups.

There are many genetic testing including different number of genes to explore the candidate gene in different cancer families, and the positive results are usually not high, which may result from ethnic, population difference and gene numbers including type of genes.³⁵ In this study, in the non-familial cancer groups, there are over 6% having P variants of cancer-related genes, most of them are the causes of familial cancer and high penetrance. Therefore, WGS should be used for screening for hereditary cancer to avoid false negative finding or de novo mutation. In addition, the number of known hereditary cancer genes has increased and it has been demonstrated that germline susceptibility to cancer is more prevalent than formerly believed; our findings also confirm these concept.^{36,37}

We used prediction tool CADD as a major tool to analyze the null variants on ClinVar database, this tool is a popularly used evaluator of variant deleteriousness

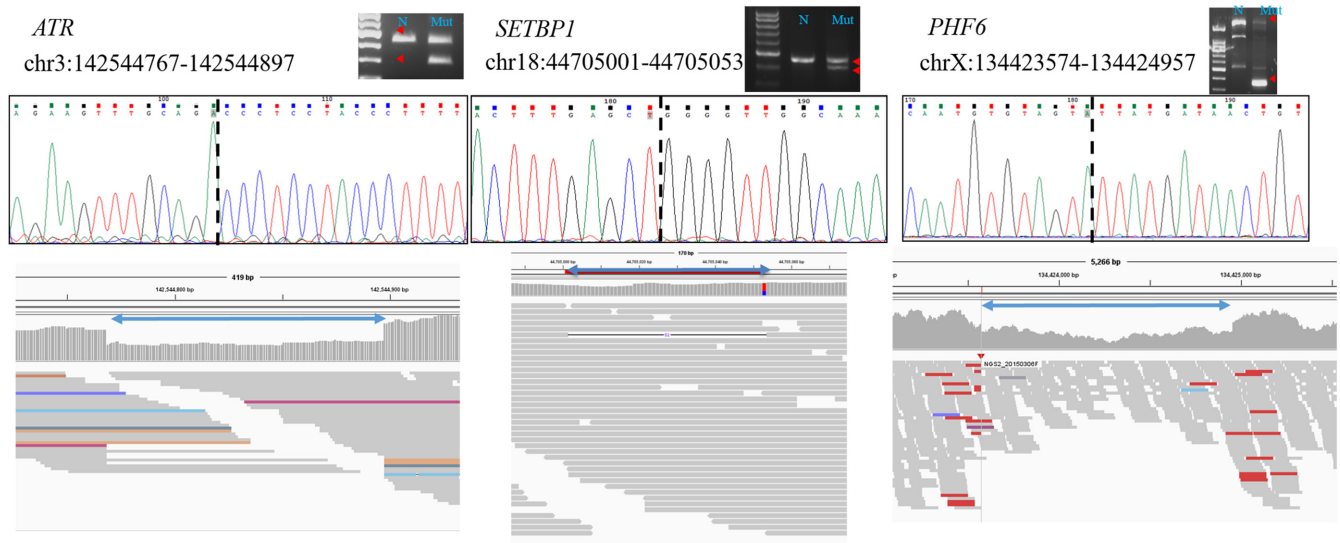


FIGURE 4 Sanger sequencing of structural variants.

that can effectively and efficiently cataloged causal variants in genetic analyses, especially for highly penetrants of severe Mendelian disorders. CADD has integrated more than 60 genomic features to annotate the variants, and can score human single nucleotide variants and short insertion and deletions anywhere in the reference assembly.^{29,30} We used CADD score ≥ 30 and MAF $< 0.5\%$ as a variant selection standard for null variants in ClinVar database of coding areas of protein-coding genes, or CADD ≥ 30 and MAF $< 0.5\%$ and FunSeq2 ≥ 0.5 and RegulomeDB ≥ 0.5 as a variant selection standard for null variants in ClinVar database of non-coding areas of protein-coding genes, or CADD score ≥ 25 for non-coding driver genes and regulatory areas as an important P variants for cancer-related genes. Using these strict criteria, we were able to avoid over-presented cancer related variants, but we may lose some false negative variants. There are over 40 different in silico programs for variant prediction, CADD is only one of them. In addition to CADD, we also used RegulomeDB, and FunSeq2 to improve the prediction accuracy.

Driver-gene screening may include many low- to moderate-penetrance genes, but the optimal strategy for managing carriers of low to moderate-penetrance mutations is unclear. Many variants of unknown significance will be detected using WGS. More data on cases with driver gene mutations is needed to evaluate their clinical significance and prevent overdiagnosis.^{38,39}

WGS enables detection of P and LP variants in cancer-related genes to identify cancer carriers, and WGS is rapid and more cost-effective than other methods used for cancer carrier screening, such as WES.⁴⁰

AUTHOR CONTRIBUTIONS

Conceptualization: Ya-Sian Chang, Jan-Gowth Chang; Data curation: Hou-Wei Chu, Shun-Jen Chang, Chieh-Min Chang; Formal analysis: Dy-San Chao, Yu-Pao Chou, Hon-Da Chen, Ya-Sian Chang, Jan-Gowth Chang; Funding acquisition: Jan-Gowth Chang; Methodology: Dy-San Chao, Chin-Chun Chung, Yu-Pao Chou, Chia-Li Lin, Hon-Da Chen, Ting-Yuan Liu; Project administration: Center for Precision Medicine, CMUH; Resources: CMUH, and TWB; Software: CADD, REVEL, RegulomeDB, and FunSeq2; Supervision: Jan-Gowth Chang; Validation: Ya-Sian Chang, Chia-Li Lin; Visualization: Yu-Hsuan Juan; Writing—original draft: Ya-Sian Chang, Jan-Gowth Chang; Writing—review & editing: Ya-Sian Chang, Jan-Gowth Chang

ACKNOWLEDGMENTS

This work was supported by grant (DMR-107-099) from the China Medical University Hospital.

CONFLICT OF INTEREST

None.

DATA AVAILABILITY STATEMENT

Raw data were generated from Taiwan Biobank (https://www.twbiobank.org.tw/new_web_en/about-export.php). The datasets generated during and/or analyzed during the current study are not publicly available due to the data containing sensitive information about patients but are available from the corresponding author upon reasonable request.

ETHICAL APPROVAL STATEMENT

The study was conducted with the approval of the Institutional Review Board of China Medical University

Hospital (CMUH 108-REC1-091), of the IRB on Biomedical Science Research/IRB-BM Academia Sinica, Taiwan (TWBR-02). The written informed consent has been obtained from all the patients included in this study.

ORCID

Jan-Gowth Chang  <https://orcid.org/0000-0003-0375-1427>

REFERENCES

- GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2020;396(10258):1204–1222. doi:10.1016/S0140-6736(20)30925-9
- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249. doi:10.3322/caac.21660
- Foulkes WD. Inherited susceptibility to common cancers. *N Engl J Med*. 2008;359(20):2143–2153. doi:10.1056/NEJMra0802968
- Samadder NJ, Giridhar KV, Baffy N, Riegert-Johnson D, Couch FJ. Hereditary cancer syndromes—a primer on diagnosis and management: part 1: breast-ovarian cancer syndromes. *Mayo Clin Proc*. 2019;94(6):1084–1098. doi:10.1016/j.mayocp.2019.02.017
- Samadder NJ, Baffy N, Giridhar KV, Couch FJ, Riegert-Johnson D. Hereditary cancer syndromes—a primer on diagnosis and management, part 2: gastrointestinal cancer syndromes. *Mayo Clin Proc*. 2019;94(6):1099–1116. doi:10.1016/j.mayocp.2019.01.042
- Zhang J, Walsh MF, Wu G, et al. Germline mutations in predisposition genes in pediatric cancer. *N Engl J Med*. 2015;373(24):2336–2346. doi:10.1056/NEJMoa1508054
- Nagy R, Sweet K, Eng C. Highly penetrant hereditary cancer syndromes. *Oncogene*. 2004;23(38):6445–6470. doi:10.1038/sj.onc.1207714
- Tung N, Domchek SM, Stadler Z, et al. Counselling framework for moderate-penetrance cancer-susceptibility mutations. *Nat Rev Clin Oncol*. 2016;13(9):581–588. doi:10.1038/nrcli.nonc.2016.90
- Wu S, Zhu W, Thompson P, Hannun YA. Evaluating intrinsic and non-intrinsic cancer risk factors. *Nat Commun*. 2018;9(1):3490. doi:10.1038/s41467-018-05467-z
- Fanfani V, Citi L, Harris AL, Pezzella F, Stracquadanio G. The landscape of the heritable cancer genome. *Cancer Res*. 2021;81(10):2588–2599. doi:10.1158/0008-5472.CAN-20-3348
- Hu C, Hart SN, Gnanaolivu R, et al. A population-based study of genes previously implicated in breast cancer. *N Engl J Med*. 2021;384(5):440–451. doi:10.1056/NEJMoa2005936
- Yang X, Song H, Leslie G, et al. Ovarian and breast cancer risks associated with pathogenic variants in RAD51C and RAD51D. *J Natl Cancer Inst*. 2020;112(12):1242–1250. doi:10.1093/jnci/djaa030
- Breast Cancer Association C, Dorling L, Carvalho S, Allen J, et al. Breast cancer risk genes – association analysis in more than 113,000 women. *N Engl J Med*. 2021;384(5):428–439. doi:10.1056/NEJMoa1913948
- Kastrinos F, Samadder NJ, Burt RW. Use of family history and genetic testing to determine risk of colorectal cancer. *Gastroenterology*. 2020;158(2):389–403. doi:10.1053/j.gastro.2019.11.029
- Vietri MT, D'Elia G, Caliendo G, et al. Prevalence of mutations in BRCA and MMR genes in patients affected with hereditary endometrial cancer. *Med Oncol*. 2021;38(2):13. doi:10.1007/s12032-021-01454-5
- Janssen B, Bellis S, Koller T, Tischkowitz M, Liao SS. A systematic review of predicted pathogenic PALB2 variants: an analysis of mutational overlap between epithelial cancers. *J Hum Genet*. 2020;65(2):199–205. doi:10.1038/s10038-019-0680-7
- Nones K, Johnson J, Newell F, et al. Whole-genome sequencing reveals clinically relevant insights into the aetiology of familial breast cancers. *Ann Oncol*. 2019;30(7):1071–1079. doi:10.1093/annonc/mdz132
- Vogelstein B, Kinzler KW. The path to cancer—three strikes and you're out. *N Engl J Med*. 2015;373(20):1895–1898. doi:10.1056/NEJMp1508811
- Rahman N. Realizing the promise of cancer predisposition genes. *Nature*. 2014;505(7483):302–308. doi:10.1038/nature12981
- Huang KL, Mashl RJ, Wu Y, et al. Pathogenic germline variants in 10,389 adult cancers. *Cell*. 2018;173(2):355–370.e14. doi:10.1016/j.cell.2018.03.039
- Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. *Nature*. 2020;578(7793):82–93. doi:10.1038/s41586-020-1969-6
- Rheinbay E, Nielsen MM, Abascal F, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*. 2020;578(7793):102–111. doi:10.1038/s41586-020-1965-x
- Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*. 2018;173(2):371–385.e18. doi:10.1016/j.cell.2018.02.060
- Martínez-Jiménez F, Muiños F, Sentís I, et al. A compendium of mutational cancer driver genes. *Nat Rev Cancer*. 2020;20(10):555–572. doi:10.1038/s41568-020-0290-x
- Fan CT, Lin JC, Lee CH. Taiwan Biobank: a project aiming to aid Taiwan's transition into a biomedical island. *Pharmacogenomics*. 2008;9(2):235–246. doi:10.2217/14622416.9.2.235
- Juang JJ, Lu TP, Su MW, et al. Rare variants discovery by extensive whole-genome sequencing of the Han Chinese population in Taiwan: applications to cardiovascular medicine. *J Adv Res*. 2021;30:147–158. doi:10.1016/j.jare.2020.12.003
- Lin JC, Fan CT, Liao CC, Chen YS. Taiwan Biobank: making cross-database convergence possible in the big data era. *Gigascience*. 2018;7(1):1–4. doi:10.1093/gigascience/gix110
- Landrum MJ, Chitipiralla S, Brown GR, et al. ClinVar: improvements to accessing data. *Nucleic Acids Res*. 2020;48(D1):D835–D844. doi:10.1093/nar/gkz972
- Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886–D894. doi:10.1093/nar/gky1016
- Rentzsch P, Schubach M, Shendure J, Kircher M. CADD-splice: improving genome-wide variant effect prediction using deep

- learning-derived splice scores. *Genome Med.* 2021;13(1):31. doi:[10.1186/s13073-021-00835-9](https://doi.org/10.1186/s13073-021-00835-9)
31. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012;22(9):1790-1797. doi:[10.1101/gr.137323.112](https://doi.org/10.1101/gr.137323.112)
 32. Fu Y, Liu Z, Lou S, et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol.* 2014;15(10):480. doi:[10.1186/s13059-014-0480-5](https://doi.org/10.1186/s13059-014-0480-5)
 33. Miller DT, Lee K, Chung WK, et al. ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med.* 2021;23(8):1381-1390. doi:[10.1038/s41436-021-01172-3](https://doi.org/10.1038/s41436-021-01172-3)
 34. Chang YS, Tu SJ, Chen YC, et al. Mutation profile of non-small cell lung cancer revealed by next generation sequencing. *Respir Res.* 2021;22(1):3. doi:[10.1186/s12931-020-01608-5](https://doi.org/10.1186/s12931-020-01608-5)
 35. Jing L, Su L, Ring BZ. Ethnic background and genetic variation in the evaluation of cancer risk: a systematic review. *PLoS One.* 2014;9(6):e97522. doi:[10.1371/journal.pone.0097522](https://doi.org/10.1371/journal.pone.0097522)
 36. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med.* 2000;343(2):78-85. doi:[10.1056/nejm200007133430201](https://doi.org/10.1056/nejm200007133430201)
 37. Kuhlen M, Taeubner J, Brozou T, Wieczorek D, Siebert R, Borkhardt A. Family-based germline sequencing in children with cancer. *Oncogene.* 2019;38(9):1367-1380. doi:[10.1038/s41388-018-0520-9](https://doi.org/10.1038/s41388-018-0520-9)
 38. Srivastava S, Koay EJ, Borowsky AD, et al. Cancer overdiagnosis: a biological challenge and clinical dilemma. *Nat Rev Cancer.* 2019;19(6):349-358. doi:[10.1038/s41568-019-0142-8](https://doi.org/10.1038/s41568-019-0142-8)
 39. Loomans-Kropp HA, Umar A. Cancer prevention and screening: the next step in the era of precision medicine. *NPJ Precis Oncol.* 2019;19:3. doi:[10.1038/s41698-018-0075-9](https://doi.org/10.1038/s41698-018-0075-9)
 40. Guzauskas GF, Garbett S, Zhou Z, et al. Cost-effectiveness of population-wide genomic screening for hereditary breast and ovarian cancer in the United States. *JAMA Netw Open.* 2020;3(10):e2022874. doi:[10.1001/jamanetworkopen.2020.22874](https://doi.org/10.1001/jamanetworkopen.2020.22874)

SUPPORTING INFORMATION

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

How to cite this article: Chang Y-S, Chao D-S, Chung C-C, et al. Cancer carrier screening in the general population using whole-genome sequencing. *Cancer Med.* 2023;12:1972-1983. doi: [10.1002/cam4.5034](https://doi.org/10.1002/cam4.5034)