

# SCIENTIFIC REPORTS

There are amendments to this paper

OPEN

## Germline susceptibility variants impact clinical outcome and therapeutic strategies for stage III colorectal cancer

Peng-Chan Lin<sup>1,4</sup>, Yu-Min Yeh<sup>1,2</sup>, Pei-Ying Wu<sup>3</sup>, Keng-Fu Hsu<sup>3</sup>, Jang-Yang Chang<sup>1,5</sup> & Meng-Ru Shen<sup>3,6</sup>

Although somatic mutations are the main cause of cancer, underlying germline alterations may affect cancer outcome. There is little information on comprehensive analysis of germline genome sequencing for cancer healthcare strategy. Here we studied the implication of germline cancer-associated variants on cancer counselling and therapeutic strategies by germline whole genome and tumor targeted sequencing. Fifty-five gynecological and 104 colorectal cancer (CRC) patients were enrolled. We identified 22 germline pathogenic variants in 16 cancer-associated genes. Most of them are involved in DNA repair signaling, including *MLH1*, *BRCA1/2*, *MUTYH*, *ATM*, *PMS2*, *MSH6*, *BAP1*, and *FANCA*. About 6% of cancer patients presented the secondary findings of germline variants with non-oncogenic impact, mainly on the cardiovascular system which should be carefully monitored during chemotherapy. CRC patients carrying germline susceptibility variants had better disease-free survival than those without variants. Importantly, in the CRC model, the underlying germline alterations mold the tumor somatic alteration landscape. *NOTCH1* mutation was the most common somatic mutation in recurrent CRC, implying a potential therapeutic target in adjuvant setting. In conclusion, both tumor genome and germline sequence data have to be analyzed to have a more complete picture of the overall genetic foundation of cancer.

Genomic medicine has been widely applied in the diagnosis, treatment, and prevention of cancer. The identification of critical genetic mutations is an important diagnostic criterion for certain types of cancer, such as the *BCR-ABL* fusion gene in chronic myeloid leukemia<sup>1</sup>. In cancer treatment, detecting recurrent oncogenic mutations has become standard practice. Currently, *EGFR*, *ALK*, and *ROS1* mutations in non-small cell lung cancer<sup>2</sup>, *RAS* and *BRAF* mutations in colorectal cancer<sup>3</sup>, and *BRAF* mutations in melanoma are routinely used in clinical practice<sup>4</sup>. The individual differences resulting in distinct treatment outcome and toxicity are becoming importance. The Clinical Pharmacogenetics Implementation Consortium (CPIC) therefore recommends analyzing patient genetic polymorphisms before chemotherapy to avoid excess toxicities, such as *UGT1A1* genotyping when prescribing irinotecan<sup>5</sup>.

Advances in next generation sequencing (NGS) have led to the identification of hundreds of mutated genes, including single nucleotide variations (SNVs), copy number variations (CNVs), small insertion/deletions (indels), and fusion genes in a single assay. Mutations in germline cancer susceptibility genes may contribute to hereditary cancer risks. Targeted multigene panels for testing inherited cancer susceptibility are also applied in the diagnosis of hereditary cancer predisposition<sup>6,7</sup>. As technical barriers and genome sequencing costs decrease, whole genome sequencing (WGS) and whole exome sequencing (WES) are increasingly used to establish the

<sup>1</sup>Department of Internal Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan. <sup>2</sup>Graduate Institute of Clinical Medicine, National Cheng Kung University, Tainan, Taiwan. <sup>3</sup>Department of Obstetrics and Gynecology, National Cheng Kung University, Tainan, Taiwan. <sup>4</sup>Department of Computer Science and Information Engineering, College of Electrical Engineering and Computer Science, National Cheng Kung University, Tainan, Taiwan. <sup>5</sup>National Institute of Cancer Research, National Health Research Institutes, Tainan, Taiwan. <sup>6</sup>Department of Pharmacology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan. Correspondence and requests for materials should be addressed to M.-R.S. (email: [mrshen@mail.ncku.edu.tw](mailto:mrshen@mail.ncku.edu.tw))

Received: 20 June 2018

Accepted: 19 February 2019

Published online: 08 March 2019

mutation landscape of diseases<sup>8,9</sup>. However, there is a lack of comprehensive analysis of germline genetic variants associated with increased cancer susceptibility combined with different treatment responses and drug-related toxicity. Moreover, germline WGS can detect pathogenic variants in the secondary finding genes recommended by the American College of Medical Genetics and Genomes (ACMG), which likely have an impact during cancer treatment.

For patients with a family history of cancer or genetic predisposition to cancer, appropriate risk assessment, accurate detection of causal genes, and proper cancer treatment and surveillance are all crucial, not only for patients themselves but also for their families. Cancer susceptibility genes may affect different molecular signal pathways in cancer formation which impact distinct responses to treatments<sup>10</sup>. Various targeted cancer panels have been developed and clinically applied to identify somatic mutations in genes or pathways that can be targeted therapeutically<sup>11,12</sup>. However, clinically favorable responses to actionable mutations with matched therapy are limited because there are few FDA-approved companion therapies, ambiguous scientific contexts, and loose treatment algorithms<sup>13</sup>. Reviewing these clinical trials<sup>14,15</sup> with a molecular profiling approach indicates that the treatment algorithms for selecting matched molecularly-targeted agents might be inadequate, and knowing the underlying mechanism of biological pathways may be beneficial to the outcome of advanced or refractory cancer patients.

Here we utilized the germline WGS and tumor deep targeted sequencing to identify germline susceptibility variants that may impact the clinical outcome and therapeutic strategies. We also highlighted the secondary genomic findings on non-oncogenic phenotypes associated with life-threatening adverse events during cancer treatment.

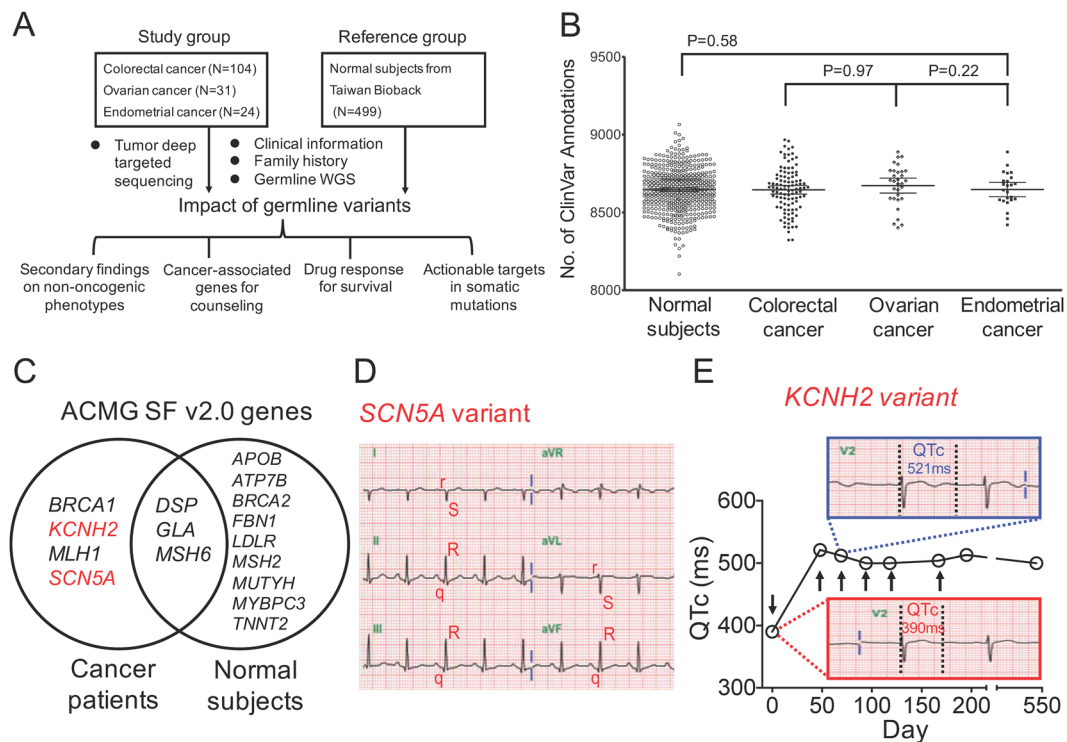
## Materials and Methods

**Patient and normal healthy subject enrollment and counseling.** All methods of this study were performed in strict accordance with the relevant guidelines and regulations. The National Cheng Kung University Hospital institutional review board approved this study (A-ER-103-395 and A-ER-104-153), and all participants gave informed written consent. A total of 104 colorectal (CRC), 31 ovarian, and 24 endometrial cancer patients were recruited as the study group in National Cheng Kung University Hospital (NCKUH) between January 2015 and March 2017. Follow-up continued through March, 2018. The follow-up period is every three months until progression or death. (Supplementary Table 1). All CRC patients were pathological stage III and received standard surgical resection followed by adjuvant chemotherapy with the regimen of mFOLFOX6 (5-fluorouracil, leucovorin, and oxaliplatin). For endometrial or ovarian cancer, carboplatin plus paclitaxel were used as post-operative adjuvant chemotherapy. Clinical information, including detailed cancer family history and blood sampling for WGS, was collected at the time of enrollment. WGS, health, and lifestyle data of 499 non-cancer normal Taiwanese people were obtained from Taiwan Biobank as a reference group.

**Germline whole genome sequencing.** Whole blood was collected for genomic DNA extraction. Genomic DNA was quantified with a Qubit fluorescence assay (Thermo Fisher Scientific) and sheared with a S2 instrument (Covaris). Library preparation was carried out using the TruSeq DNA PCR-Free HT kit (Illumina). Individual DNA libraries were measured by 2100 Bioanalyzer (Agilent) qPCR and Qubit (Thermo Fisher Scientific). Normalized DNA libraries were combined into five-sample pools per flow cell in all eight lines and clustered on a cBot instrument (Illumina) with Paired-End Cluster Kit V4 (Illumina). All flow cells were sequenced on the HiSeq. 2500 sequencer (Illumina) using SBS kit V4 chemistry (Illumina). FastQC was used to check read quality, and the resulting reads were aligned to the hg19 reference genome with the BWA-MEM algorithm<sup>16</sup>. SNV and indel identification and genotyping were performed across all samples simultaneously using standard hard filtering parameters or variant quality score recalibration according to GATK Best Practices recommendations<sup>17,18</sup>. WGS was performed with a minimum median coverage of 30X.

**Targeted tumor sequencing by cancer panel.** A total of 107 primary tumor samples were sent for histologic assessment followed by extraction of nucleic acids from formalin-fixed paraffin-embedded blocks in NCKUH. Specimens were reviewed by pathologists who determined the percentage of viable tumor nuclei and adequacy for profiling mutation detection. Tumor deep targeted sequencing was performed by OncoPrint Comprehensive Assays (OCA) (Thermo Fisher Scientific)<sup>19</sup>. The OCA was designed to detect 143 solid tumor-related genes, including 73 hotspot genes, 49 focal CNV gains, 26 genes for full coding region sequencing (CDS), and 22 fusion driver genes. The Ion PGM Sequencing 200 Kit v.2 was used with the Ion PGM sequencer (Thermo Fisher Scientific) according to the manufacturer's instructions. All samples were analyzed using the Torrent Suite Software 5.0.4, aligning all reads to the hg19 reference genome, and variant calling was performed running the Torrent Variant Caller plugin version 5.0.4.0.

**Genetic susceptibility variants selected for analysis.** We used the ANNOVAR<sup>20</sup> tool to annotate SNVs and indels and filter out variants reported in the 1000 Genomes Project, the Single Nucleotide Polymorphism Database (dbSNP), and allele frequency cutoffs above 0.1% in the Exome Aggregation Consortium (ExAC) database<sup>21</sup>. We chose definitions to select for germline genetic susceptibility variants, including germline cancer-associated genes, drug response genes, and ACMG secondary finding genes. Cancer-associated genetic variants classified as pathogenic variants according to the review of experts were identified and then filtered by our in-house control, the database from Taiwan Biobank. Clinical significance and classification criteria of genetic variants are defined by National Center for Biotechnology Information (NCBI)-ClinVar, a freely available curated database of relationships among medically important genetic variants and phenotypes<sup>22,23</sup>. The 565 well-known cancer-associated genes reviewed by a previous study<sup>24</sup> and genetic variants associated with the cancer phenotype summarized in the ClinVar database were selected for analysis. The Pharmacogenomics Knowledgebase



**Figure 1.** Study flow chart and the genomic findings on potential life-threatening adverse events during chemotherapy. **(A)** Flowchart of WGS and tumor deep targeted sequencing analyses for the identification of germline genetic variants and their implications on patient outcomes. A total of 159 cancer patients and 499 normal subjects were enrolled. WGS data generated from germline DNA were used to analyze the presence of single-nucleotide variants (SNVs) and small insertions and deletions (indels). Identified genetic variants were correlated to clinical information to determine the impacts of germline variants. **(B)** Numbers of ClinVar annotation in the normal population, colorectal, ovarian, and endometrial cancer patients. Each genetic variant was compared with published alterations in the ClinVar database. The number of genetic variants with ClinVar annotation was compared between the normal population and different types of cancers by *t*-test. **(C)** ACMG SF v2.0 genes with pathogenic variants in 159 cancer patients and 499 normal subjects. Germline genetic variants were compared with the list of genes recommended by ACMG to be reported as incidental or secondary findings (SFs). ACMG SF v2.0 genes with pathogenic variants detected in cancer patients and normal subjects were listed in the left and right circle, respectively. **(D)** The EKG shows right axis deviation, small R waves with deep S waves (rS complexes) in leads I, and aVL and small Q waves with tall R waves (qR complexes) in II, III, and aVF, suggesting left posterior fascicular block. **(E)** Corrected QT (QTc) interval in an endometrial cancer patient at baseline and after the first cycle of treatment with carboplatin and paclitaxel. Changes in QTc interval before, during, and after the six cycles of chemotherapy. EKG was performed at baseline (before surgery), before each cycle of carboplatin plus paclitaxel, and one year after chemotherapy. Arrows indicate the timing of surgery and chemotherapy, respectively.

(PharmGKB) provides information about the impact of human genetic variation on drug responses<sup>25</sup> and is used for selecting drug response and adverse drug reaction genetic variants found among chemotherapy-related genetic variants. The updated ACMG secondary findings recommendations were used to identify genetic variants with important non-oncogenic phenotypes<sup>26</sup>. The analytic workflow of drug response and ACMG genetic variants includes automatic extraction matching and database annotation (Supplementary Fig. 1).

**Statistical analysis.** Chi-Square test, Fisher's exact test, and unpaired *t*-test were used to assess the difference between groups. Kaplan–Meier curves were used to assess disease-free survival. Disease-free survival was defined as the time between surgery and recurrence of cancer. A *P*-value < 0.05 was considered statistically significant.

## Results

**Study flow chart and genomic findings on potential life-threatening adverse events.** Figure 1A shows the study design, in which 159 cancer patients and 499 normal healthy subjects were enrolled as the study and reference groups, respectively. In the study group, 104 (65%) patients were CRC and 55 (35%) patients were gynecological cancer (Supplementary Table 1). Forty percentage of cancer patients presented the family history of cancer in the first-degree relatives, and 12% of them showed cancer family history in the second-degree relatives. In terms of ClinVar annotation, 1,369,942 genetic variants (SNVs and small indels) were annotated in 159 cancer patients. A total of 48,419 genetic variants (reviewed by experts), including pathogenic, uncertain significance,

Gene	Name	Cancer patients (%)	Normal subjects (%)	dbGaP (%)
FGFR4	FGFR4:c.1162G > A (p.Gly388Arg)	71.07	69.14	45.8
PTPRJ	PTPRJ:c.827A > C (p.Gln276Pro)	49.06	48.70	26.9
PRSS1	PRSS1:c.161A > G (p.Asn54Ser)	29.56	32.46	0.42
SLC22A18	SLC22A18:c.257G > A (p.Arg86His)	16.35	12.42	1.35
PRSS1	PRSS1:c.47C > T (p.Ala16Val)	13.21	10.62	7.18
EHBP1	EHBP1:c.1185 + 30064G > A	11.32	5.81	NA
RAD54B	RAD54B:c.1778A > G (p.Asn593Ser)	1.26	3.01	0.86
MLH1	MLH1:c.649C > T (p.Arg217Cys)	1.26	1.00	NA
TEK	TEK:c.2228G > C (p.Gly743Ala)	0.63	1.00	NA
MSH6	MSH6:c.4001 + 2_4001 + 5delTAAC	0.63	1.00	NA

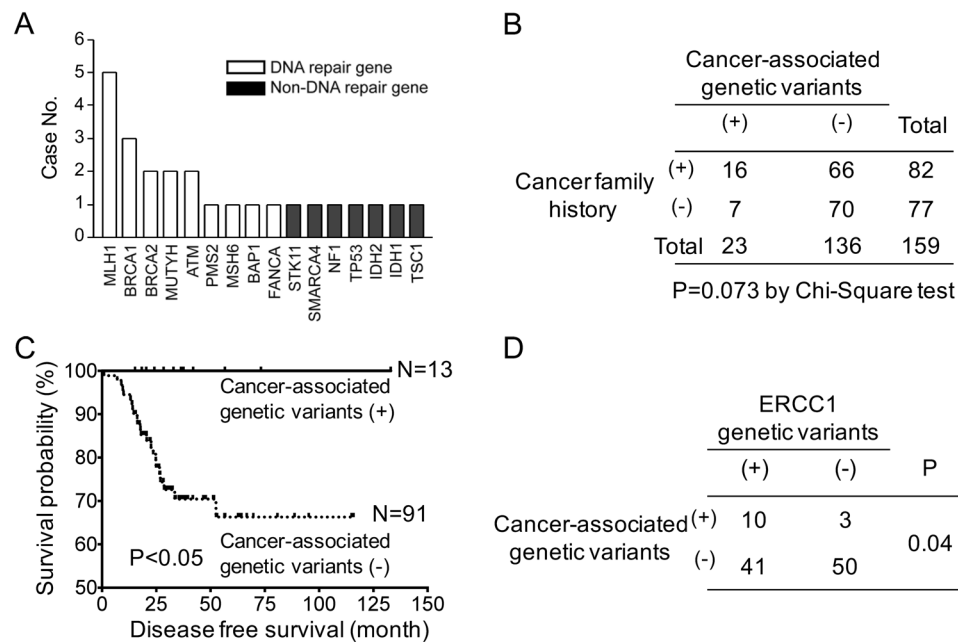
**Table 1.** In-house filtering for cancer-associated genetic variants. Based on the public dataset, candidates of germline genetic variants associated with cancer phenotypes were identified in cancer patients and normal subjects. The top 10 genetic variants with the frequency  $\geq 1\%$  detected in normal subjects were listed. N the dbGaP, the frequency of *PRSS1* c.161A > G, *SLC22A18* c.257G > A, and *RAD54B* c.1778A > G was 0.42, 1.35, and 0.86%, respectively. On the other hand, these three genetic variants had relatively high frequency in our cancer patients and normal subjects, indicating that the pathogenicity of these variants for cancer was less likely in Taiwanese population. dbGaP: the database of Genotypes and Phenotypes; NA: not available.

drug response, benign, and likely benign categories, were identified. “Reviewed by experts” indicates that the evidence to support the interpretation of genetic variants is provided by a ClinGen-approved expert panel<sup>23</sup>. There were 8,323 to 8,967 genetic variants (median: 8,661) per cancer patient and 8,104 to 9,065 genetic variants (median: 8,649) per normal subject (Fig. 1B). There was no significant difference in ClinVar annotation numbers between cancer patients and normal subjects.

Germline genetic variants associated with non-oncogenic phenotypes were first analyzed by comparing with ACMG secondary findings (SF v2.0) gene list<sup>26</sup>. We identified 12 ACMG SF v2.0 genetic variants in 23 of 499 normal subjects (4.6%), and 7 genetic variants were identified in 9 of 159 cancer patients (5.6%) (Fig. 1C, Supplementary Table 2), including *BRCA1*, *KCNH2*, *MLH1*, *SCN5A*, *DSP*, *GLA*, and *MSH6*. Four of these 7 genetic variants, including *KCNH2*, *SCN5A*, *DSP*, and *GLA*, were associated with cardiovascular diseases. For instance, *SCN5A* is associated with Brugada syndrome-1 (MIM 601144)<sup>27</sup>. The *SCN5A* variant was identified in a 56-year-old patient with ovarian cancer. This patient did not have past history of heart disease. Although no typical manifestation of Brugada syndrome was observed, the EKG demonstrated right axis deviation, rS complexes in leads I and aVL, and qR complexes in leads II, III, and aVF, which suggested left posterior fascicular block (Fig. 1D). *KCNH2* is associated with long QT syndrome 2 (MIM 613688)<sup>27</sup>. One 63-year-old patient with endometrial cancer had the c.1714G > A (p.Gly572Ser) variant in *KCNH2*. This patient received the standard surgical resection followed by adjuvant chemotherapy with the regimen of carboplatin and paclitaxel for 6 cycles. The baseline EKG, which was performed before surgery, showed normal sinus rhythm with the QTc around 390 ms (Fig. 1E). After the first cycle of chemotherapy, the QTc interval significantly increased from 390 to 521 ms (Fig. 1E). This patient received a total of 6 cycles of chemotherapy. The QTc interval was around 500–515 ms during the period of chemotherapy and the prolonged QTc was still noted 1 year after the completion of chemotherapy (Fig. 1E). We have reviewed the family history for the individuals with “incidental findings”. There are no associated diseases with “incidental findings” in the first-degree and the second-degree relatives. These findings highlight the importance of identifying germline variants associated with non-oncogenic phenotypes which should be carefully monitored during cancer treatment.

**Ethnic difference of germline susceptibility variants.** Identifying germline mutations in cancer susceptibility genes was important for genetic counseling. The ethnic difference in the genomic background should be taken into consideration to minimize false-positive results. Based on the public dataset<sup>20</sup>, 46 candidates of germline genetic variants associated with cancer phenotypes were identified in the study and reference group (Supplementary Table 3). The top 10 genetic variants with the frequency  $\geq 1\%$  detected in the reference group were listed in Table 1. In the database of Genotypes and Phenotypes (dbGaP)<sup>28</sup>, the frequency of *PRSS1* c.161A > G, *SLC22A18* c.257G > A, and *RAD54B* c.1778A > G was 0.42, 1.35, and 0.86%, respectively. Contradictorily, these three genetic variants had relatively high frequency in this study (Table 1), indicating that the pathogenicity of these variants for cancer was less likely in Taiwanese population. Therefore, these three genetic variants could be misclassified as pathogenic variants in ClinVar database.

**Identification of cancer-associated genetic variants for genetic counseling.** After filtering out in-house control, 22 pathogenic variants in 16 cancer-associated genes were detected in 23 of 159 patients, including 13 (12.5%) CRC patients and 10 (18.2%) gynecologic cancer patients (Supplementary Table 4). Five patients had pathogenic variants in *MLH1*, which was the most common gene harboring a germline mutation in the study group. Other pathogenic variants were in *BRCA1*, *BRCA2*, *MUTYH*, *ATM*, *PMS2*, *MSH6*, *BAP1*, and *FANCA*, which were DNA repair-associated genes (Fig. 2A). We also detected the pathogenic variants in *STK11*, *SMARCA4*, *NF1*, *TP53*, *IDH1*, *IDH2*, and *TSC*. These genes were non-DNA repair-associated genes, and



**Figure 2.** The impact of germline cancer-associated variants on clinical outcome. **(A)** Germline cancer-associated variants identified in 55 gynecological and 104 CRC patients. Variants of cancer-associated genes in the ClinVar database were selected for analysis. After filtering out in-house control, 16 germline cancer-associated variants were detected in 23 patients. **(B)** The relationship between germline cancer-associated variants and cancer family history tested by Chi-square test. **(C)** The Kaplan-Meier curves of disease-free survival in CRC patients with and without germline cancer-associated variants. **(D)** The relationship between germline cancer-associated variants and drug response *ERCC1* gene tested by Chi-square test.

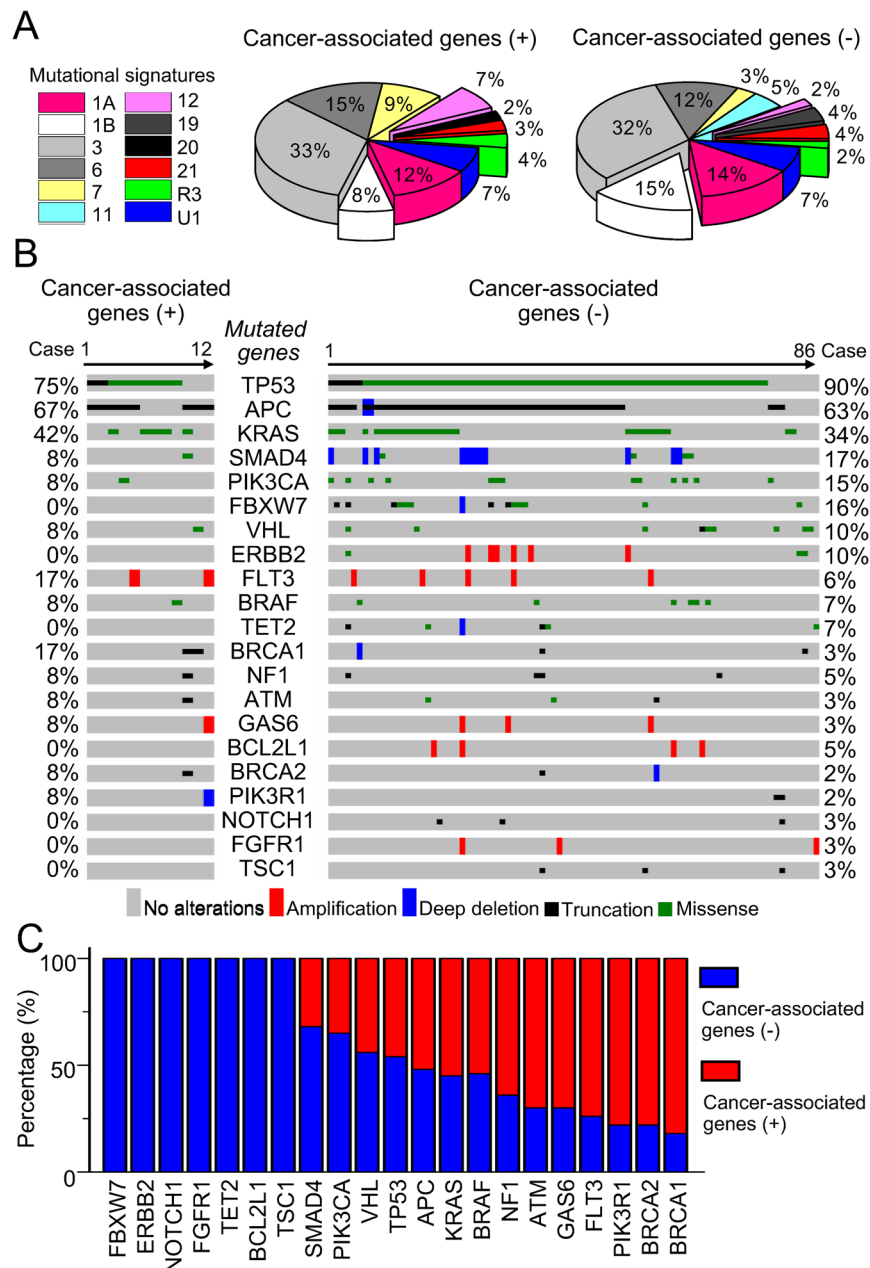
the frequency of pathogenic variants in these genes was relatively low. These results suggested that the defects in DNA repair pathways contribute to carcinogenesis and are potentially attractive targets for therapy, such as PARP inhibitors<sup>29</sup>.

We studied the association between cancer family history and germline mutations of cancer-associated genes. As shown in Fig. 2B, the incidence of germline cancer-associated genetic variants is 19.5% for patients with cancer family history, whereas the incidence is decreased to 9.1% for patients without history. There is a trend showing the correlation between the germline cancer-associated genetic variants and cancer family history in spite of non-statistical significance.

The hereditary cancer syndromes were diagnosed in three patients. Two patients harboring the *MLH1* mutation had the diagnosis of Lynch syndrome (Supplementary Fig. 2A,B). One patient had the germline variant *BRCA1* c.122A > G confirmed by Sanger sequencing. Her mother was diagnosed with ovarian cancer at the age of 50 (Supplementary Fig. 2C). However, among 23 patients with germline cancer-associated genetic variants, 7 patients did not show the family history of cancer (Fig. 2B and Supplementary Fig. 2D). Although we did not detect any germline cancer-associated variants in the other 136 patients, 66 of these patients (49%) still had a family history of cancer (Fig. 2B and Supplementary Fig. 2E).

**Impact of germline susceptibility variants on survival.** The association between cancer susceptibility variants and clinical outcome is uncertain. Here, we observed that CRC patients carrying cancer-associated genetic variants had better disease-free survival than those without variants ( $P < 0.05$ , Fig. 2C). All CRC patients in the study group were pathological stage III and received adjuvant chemotherapy with mFOLFOX6. Previous studies showed that the excision repair cross-complementing (*ERCC1*) variants were a predictive marker of oxaliplatin-based chemotherapy<sup>30</sup>. Accordingly, we studied the incidence of *ERCC1* genetic variants in CRC cancer. As shown in Fig. 2D, 77% of CRC patients with cancer-associated variants had the *ERCC1* c.354T > C variant. In contrast, 45% of CRC patients without cancer-associated variants had *ERCC1* variant ( $P = 0.04$ ). The higher percentage of *ERCC1* variant in CRC patients with pathogenic variants of cancer-associated genes might explain the better survival in this group of patients.

**Actionable targets of somatic mutations.** Compared to patients with cancer-associated genetic variants, the disease-free survival of patients without variants was shorter (Fig. 2C). We hypothesized the underlying germline alterations may mold the somatic alteration landscape and cooperate with acquired mutations to promote tumor onset and maintenance. To identify actionable somatic mutations perhaps provides opportunities to improve the clinical outcome of these patients. Therefore, we performed deep targeted sequencing analysis of 98 available CRC primary tumor samples. Twelve types of tumor mutational signatures<sup>31</sup> were observed in 98 CRC patients (Fig. 3 and Supplementary Fig. 3). The tumor mutational signature 1A, 3, 6, 7, 21, R3 and U1



**Figure 3.** Underlying germline cancer-associated variants mold the somatic alteration landscape. **(A)** Distribution of mutational signatures in CRC patients with or without cancer-associated genetic variants. Different colors indicate different types of mutational signatures. **(B)** The genomic landscapes of tumor somatic mutations in CRC patients with ( $n = 12$ ) or without ( $n = 86$ ) germline cancer-associated genetic variants. **(C)** Dominant somatic mutations in different germline backgrounds. Somatic mutations in Notch signal pathways (*FBXW7* and *NOTCH1*) or receptor tyrosine kinase signaling (*ERBB2*, *FGFR1*) exclusively happened in CRC patients without germline cancer-associated variants. Somatic mutations in DNA repair signaling (*ATM*, *BRCA1*, *BRCA2*) were mainly noted in CRC patients carrying germline cancer-associated variants.

contributed similarly between patients with or without germline cancer-associated variants. In contrast, the frequency of mutational signature 1B, 11, 12 and 19 was significantly different between two groups of patients. Signature 1B, which is an age-associated mutation signature, accounting for 8% and 15% of patients with and without germline cancer-associated variants, respectively. Moreover, tumor mutational signature 11 and 19 were exclusively observed in patients without carrying germline cancer-associated variants and mutational signature 20 could only be found in patients carrying germline cancer-associated variants. These results suggest that the underlying germline alterations indeed mold the types of tumor mutational signature.

Moreover, 434 somatic mutations in 48 genes were detected in 98 primary tumor samples (Fig. 3B). *TP53*, *APC* and *KRAS* were the most frequently mutated genes detected in patients with and without germline cancer-associated variants. In contrast, *FLT3*, *BRCA1*, *BRCA2* and *PIK3R1* mutations were more frequently

Patient	Total exonic mutations (Primary/recurrent)	Recurrent-specific somatic mutation genes
1	19/20	NOTCH1, RAC1, VHL
2	25/23	RB1, MLH1, NOTCH1, TSC1
3	33/23	PTCH1, NF1
4	19/20	NOTCH1
5	14/18	ERBB3, DNMT3A, NOTCH1
6	23/21	DNMT3A, CDH1
7	22/21	DNMT3A, NOTCH1, PIK3R1
8	22/20	NOTCH1
9	19/17	NOTCH1

**Table 2.** Protein-coding mutated genes identified in primary–recurrent tumor pairs. The somatic mutations of primary and recurrent tumor samples from 9 CRC patients were analyzed. By comparing the somatic mutations between primary and recurrent tumors, the mutated genes only detected in recurrent tumor, including NOTCH1, RAC1, MLH1, TSC1, PTCH1, NF1, ERBB3, DNMT3A, CDH1, and PIK3R, were identified.

detected in patients carrying germline cancer-associated variants (Fig. 3B,C). On the other hand, the frequency of *SMAD4* and *PIK3CA* mutation was higher in patients without germline cancer-associated variants (17% and 15%, respectively) than those with variants (8% and 8%, respectively). Moreover, mutations of *FBWX7*, *ERBB2*, *TET2*, *BCL2L1*, *NOTCH1*, *FGFR1* and *TSC1* were only detected in patients without germline cancer-associated variants (Fig. 3B,C). Mutations of *FBWX7* and *NOTCH1* lead to aberrant activation of Notch signaling pathway. *ERBB2* and *FGFR1* are well-known receptor tyrosine kinases involved in the carcinogenesis of different types of cancers. These results imply that patients with and without germline cancer-associated variants may harbor different tumor somatic mutations. Notch, *ERBB2*, and *FGFR1* signaling pathway might be potential therapeutic targets to improve the outcome of CRC patients without germline cancer-associated genetic variants in adjuvant setting.

**Emerging NOTCH1 mutated subclones in tumor recurrence.** To identify potential therapeutic targets, we analyzed the somatic mutations of primary and recurrent tumor samples from 9 CRC patients with recurrent diseases. By comparing the somatic mutations between primary and recurrent tumors, we identified the mutated genes only detected in recurrent tumor, including *NOTCH1*, *RAC1*, *MLH1*, *TSC1*, *PTCH1*, *NF1*, *ERBB3*, *DNMT3A*, *CDH1*, and *PIK3R* (Table 2). The most common mutated gene detected in recurrent tissue was *NOTCH1* gene, which was identified in 7 of 9 CRC patients.

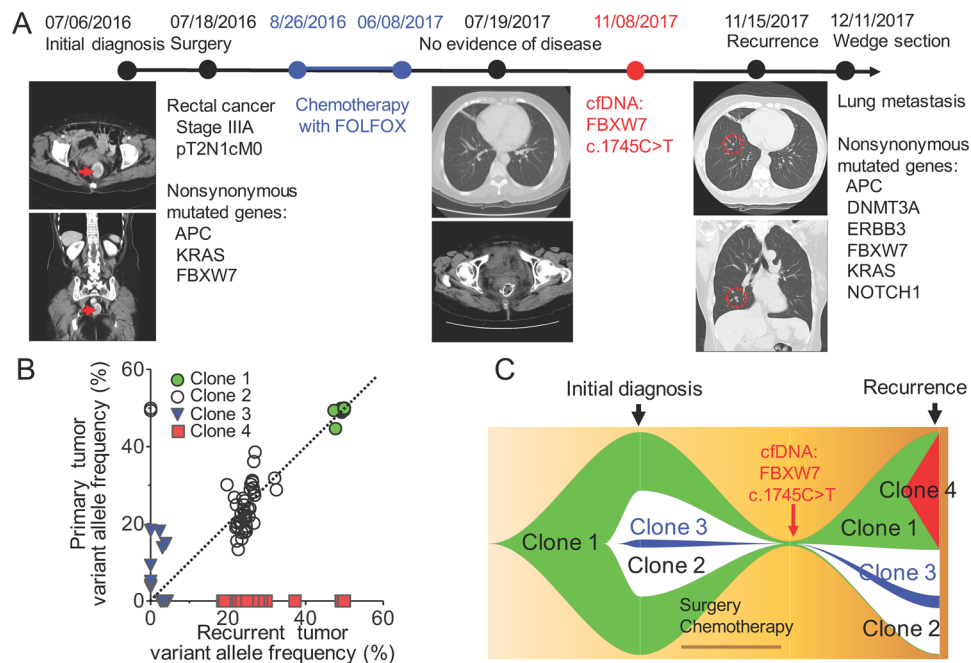
Figure 4 illustrates a case study showing that Notch signaling pathway might be an important therapeutic target for stage III CRC patients in adjuvant setting. This case is the patient number 5 (Table 2 and Supplementary Table 5), a 66-year-old woman with pathological stage III rectal cancer at initial diagnosis. She did not carry any germline cancer-associated variants and tumor targeted gene sequencing showed somatic mutations in *APC*, *KRAS*, and *FBWX7* in the primary tumor sample. She received standard surgical resection followed by adjuvant chemotherapy with mFOLFOX6 for 12 cycles. No disease recurrence was detected by computed tomography (CT) scan 12 months after surgery. However, when cell-free circulating DNA (cfDNA) was isolated 16 months after surgery, analysis of tumor-specific mutations showed the detection of *FBWX7* c.1745C > T mutation and subsequent CT scan confirmed the recurrence with pulmonary metastasis. Wedge resection of pulmonary metastasis was performed under the curative intent. Three new somatic mutations, including *ERBB3*, *DNMT3A* and *NOTCH1* mutations were detected in the metastatic tumor.

Understanding the clonal evolution during cancer progression offers the insight into tumor heterogeneity and potential drug targets. The variant allele frequency and clustering of variants with similar cellular prevalence were used to reconstruct the clonal evolution<sup>32</sup>. In this case, there are 3 clusters of tumor clones identified in the primary tumor (Fig. 4B,C). The clonal evolution model revealed that clone 1 was the founding clone and the clone 2 and 3 were subsequently derived from clone 1. By contrast, 4 clusters of clones were identified in the recurrent tumor tissues and the clone 3 still existed with additional *NOTCH1* mutation. The detection of *FBWX7* mutation by cfDNA during post-operative follow-up and the emerging *NOTCH1* mutation in the recurrent tumor samples highlight that Notch signaling pathway might be an important therapeutic target for stage III CRC patients in adjuvant setting.

## Discussion

With the advances of NGS technologies, germline WGS is becoming an affordable tool in health care. Here, we demonstrate the application of germline WGS in the care of cancer patients and the impact of germline genetic variants on clinical outcome and therapeutic strategies. Our results highlight the following important points: (i) Germline variants of ACMG SF genes are associated with life threatening toxicities during chemotherapy. (ii) Germline variants of cancer-associated genes affect clinical outcome and screening for the variants should take the ethnicity into consideration. (iii) Notch signaling pathway might be a therapeutic target for CRC patients without germline cancer-associated variants in adjuvant setting. These findings suggest analyzing germline variants with oncologic and non-oncologic phenotypes has a great impact on the care planning of cancer patients.

By comparing the ACMG SF v2.0 gene list, we identified 7 genetic variants with non-oncologic phenotypes in 9 of 159 cancer patients (5.6%). Most of the genetic variants were associated with cardiovascular diseases. For example, *KCNH2*, a genetic variant associated with long QT syndrome 2, was detected in an endometrial



**Figure 4.** A case study showing the clonal evolution and potential drug targets during cancer progression. **(A)** The clinical time course for a 66-year-old woman with pathological stage III rectal cancer at initial diagnosis. Serial genome sequencing and CT scan were tested. Nonsynonymous recurrent-specific mutated genes including *DNMT3A*, *ERBB3*, *NOTCH1* were detected in recurrent tissue. *FBXW7* mutations were detected in primary, recurrent tissue and cell-free DNA. **(B)** The clonal evolution during tumor progression. The variant allele frequency and clustering of variants with similar cellular prevalence were used to reconstruct the clonal evolution. **(C)** Visualizing tumor evolution with the fishplot package. The founding clone 1 (cluster 1 mutations) and the subsequent clone 2 (cluster 2 mutations) and clone 3 (cluster 3 mutations) were all detected in the recurrent tumor. Emerging *NOTCH1* variant was detected in the clone 3 of recurrent tumor tissue. *FBXW7* c.1745C>T was detected in cell-free DNA before the clinical evidence of recurrence.

cancer patient who developed QT prolongation after chemotherapy with paclitaxel. Genetic susceptibility of chemotherapy-induced cardiac toxicity has been extensively reviewed<sup>33</sup>. Based on the existing evidence, genetic testing before the initiation of treatment to predict risk and reduce the occurrence of cardiotoxicity is recommended. Our result suggests that the *KCNH2* genetic variant is a potential risk factor for paclitaxel-induced QT prolongation. More studies are needed to confirm the predictive role of *KCNH2* in chemotherapy-induced cardiac toxicity.

To identify the germline variants of cancer-associated genes and interpret the clinical significance accurately is important for both cancer patients and their family. By comparing to the pathogenic variants of cancer-associated genes in dbGaP database, we identified 3 germline variants which had relatively high frequency in cancer patients as well as normal subjects, including *PRSS1* c.161 A > G, *SLC22A18* c.257 G > A, and *RAD54B* c.1778A > G. The high frequency in normal Taiwanese population indicates the pathogenicity of these variants for cancer was less likely. This indicates that screening for cancer-associated genetic variant should take the ethnicity into consideration. Interestingly, we found that most of the cancer-associated genes harboring germline mutation were involved in DNA repair, such as *MLH1*, *BRCA1*, *BRCA2*, *MUTYH*, *ATM*, *PMS2*, *MSH6*, *BAP1*, and *FANCA*. Mutations in mismatch repair genes, including *MLH1*, *MSH2*, *MSH6*, and *PMS2*, were the well-known causative germline predisposition factors in CRCs. Recently, the detection of pathogenic variants in Fanconi anemia DNA damage repair pathway genes, such as *BRCA2*, *BRIP1*, *FANCC*, and *FANCD1* was also reported in familial CRCs<sup>34</sup>. These results show that defects in DNA repair pathways contribute to carcinogenesis of CRCs. Targeting DNA repair pathways, such as PARP inhibitors, may provide benefits for CRC patients with mutations in DNA repair genes. In this study, 12.5% of colorectal cancer (CRC) patients and 18.2% of gynecological (ovarian or endometrial) cancer patients carried pathogenic germline variants. In general, the prevalence of germline cancer susceptibility gene mutations was about 8% among cancer patients, based on a study analyzing 33 types of cancers from 10,389 cases<sup>35</sup>. However, the prevalence and spectrum of gene mutations vary in the different types of cancers. For example, of 450 patients with early-onset CRC, 16% had germline cancer susceptibility gene mutations<sup>36</sup>. In addition, 19.9% of ovarian cancer patients have been reported carrying the pathogenic germline mutation<sup>35</sup>. The impact of VUSs on genomic function was considered from different perspectives, such as cancer enrichment, LOH, protein expression and other evidence<sup>35</sup>. In this study, there were 3 patients with cancer family history carrying VUSs. These VUSs were not detected in the reference group of Taiwan BioBank. We thus considered these VUSs as potentially pathogenic variants<sup>37</sup>.

After surgical resection, adjuvant chemotherapy with mFOLFOX is the standard treatment for stage III CRC patients. The 5-year disease-free survival rate is around 73%. A lot of clinical studies have tried to further improve



the outcome of these patients. However, no progress was achieved. Here we found CRC patients without germline cancer-associated variants had a worse outcome than those with germline variants. Mutations of *FBXW7* and *NOTCH1*, which could lead to aberrant activation of Notch signaling pathway, were only detected in patients without germline cancer-associated variants. In addition, the analysis of clonal evolution also revealed *NOTCH1* mutation was the most common somatic mutation detected in recurrent tumor samples. The important roles of *NOTCH1* signaling in CRC have been well studied and it has been shown that *NOTCH1* mutations could lead to aberrant activation of Notch signaling pathway contributing cancer behaviors<sup>38,39</sup>. Here we identified *NOTCH1* mutation was the most common somatic mutation detected in recurrent CRC samples. We have checked the *NOTCH1* mutations in Integrative Genomics Viewer to rule out technical artefacts and the pathogenicity was assessed by the variant scoring systems, such as SIFT and PolyPhen<sup>40</sup>. The *NOTCH1* mutation detected in this study (p.Gly347Asp and p.His727Tyr) was predicted to affect the protein function. Further study on the biological impact of individual *NOTCH1* mutations will be evaluated in the future. These results suggest Notch signaling pathway might be a potential therapeutic target for stage III CRC patients in adjuvant setting.

Using unbiased analysis to identify the association between uncharacterized variant and cancer susceptibility or prognosis is an interesting topic that we are working on. The major aim of this study is to investigate whether germline genetic variants have an impact on clinical outcome or therapeutic strategy. That's why we comprehensively analyzed the germline genetic variants by WGS and used the established database to identify variants that are known to be implicated in cancer susceptibility, drug response and non-oncogenic phenotypes.

Despite the use of established database, our study showed several important findings. The first is the underlying germline genetic alterations could mold the tumor mutational landscape. The frequency of several tumor mutational signatures, including signature 1B, 11, 12 and 19, was significantly different between CRC patients with and without germline cancer-associated genetic variants. In addition, the somatic mutations detected in patients with and without germline cancer-associated variants were different, too. These results imply that the therapeutic strategies should be tailored by both germline background and somatic mutations. Second, the pathogenicity of several cancer-associated genetic variants, including *PRSSI* c.161A > G, *SLC22A18* c.257G > A, and *RAD54B* c.1778A > G, was misclassified. The frequency of the above genetic variants in the reference population was high (32.46, 12.42 and 3.01%, respectively). High allele frequency in population data is a strong evidence of benign impact<sup>41</sup>. Our results provide the new evidence of misclassification for these 3 germline genetic variants and suggest that the ethnic difference should be taken into consideration to minimize false-positive results. Third, 6% of cancer patients presented the secondary findings of germline variants with non-oncogenic impact, mainly on the cardiovascular system. We demonstrated abnormal cardiac electricity developed in a patient carrying the variant in *KCNH2* while receiving the adjuvant chemotherapy. The cardiac abnormality became irreversible after the completion of chemotherapy. This result suggests that patients with germline genetic variants with non-oncogenic impact should be carefully monitored during chemotherapy.

In conclusion, this study highlights the importance of applying integrative genomic information for healthcare planning and monitoring side effects during cancer treatment. Germline sequence data and tumor genome have to be analyzed together to have a more complete picture of the overall genetic foundation of cancer.

## References

1. Sabattini, E., Bacci, F., Sagramoso, C. & Pileri, S. A. WHO classification of tumours of haematopoietic and lymphoid tissues in 2008: an overview. *Pathologica* **102**, 83–87 (2010).
2. Rotow, J. & Bivona, T. G. Understanding and targeting resistance mechanisms in NSCLC. *Nat. Rev. Cancer* **17**, 637–658 (2017).
3. Zhang, J. *et al.* Molecular spectrum of KRAS, NRAS, BRAF and PIK3CA mutations in Chinese colorectal cancer patients: analysis of 1,110 cases. *Sci. Rep.* **8**, 678, 1–3 (2015).
4. Holderfield, M., Deuker, M. M., McCormick, F. & McMahon, M. Targeting RAF kinases for cancer therapy: BRAF-mutated melanoma and beyond. *Nat. Rev. Cancer* **14**, 455–467 (2014).
5. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP working group: can UGT1A1 genotyping reduce morbidity and mortality in patients with metastatic colorectal cancer treated with irinotecan? *Genet. Med.* **11**, 15–20 (2009).
6. Selkirk, C. G. *et al.* Cancer genetic testing panels for inherited cancer susceptibility: the clinical experience of a large adult genetics practice. *Fam. Cancer* **13**, 527–36 (2014).
7. LaDuca, H. *et al.* Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet. Med.* **16**, 830–7 (2014).
8. Choi, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 19096–19101 (2009).
9. Lupski, J. R. *et al.* Whole genome sequencing in a patient with charcot–marie–tooth neuropathy. *N. Engl. J. Med.* **362**, 1181–1191 (2010).
10. Alexandrov, L. B., Nik-Zainal, S., Siu, H. C., Leung, S. Y. & Stratton, M. R. A mutational signature in gastric cancer suggests therapeutic strategies. *Nat. Commun.* **29**, 8683 (2015).
11. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
12. Hyman, D. M., Taylor, B. S. & Baselga, J. Implementing genome-driven oncology. *Cell* **168**, 584–599 (2017).
13. Voest, E. E. & Bernards, R. DNA-Guided Precision Medicine for Cancer: A Case of Irrational Exuberance? *Cancer Discov.* **6**, 130–2 CD-15-1321 (2016).
14. Meric-Bernstam, F. *et al.* Feasibility of large-scale genomic testing to facilitate enrollment onto genomically matched clinical trials. *J. Clin. Oncol.* **33**, 2753–2762 (2015).
15. Le Tourneau, C. *et al.* Molecularly targeted therapy based on tumour molecular profiling versus conventional therapy for advanced cancer (SHIVA): a multicentre, open-label, proof-of-concept, randomised, controlled phase 2 trial. *Lancet. Oncol.* **16**, 1324–1334 (2015).
16. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 13033997v2 (2013).
17. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
18. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **11**, 11.10.11–11.10.33 (2013).

19. Lih, C.-J. *et al.* Analytical validation of the next-generation sequencing assay for a nationwide signal-finding clinical trial: molecular analysis for therapy choice clinical trial. *J. Mol. Diagn.* **19**, 313–327 (2017).
20. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
21. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–91 (2016).
22. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–868 (2016).
23. Rehms, H. L. *et al.* ClinGen—the Clinical Genome Resource. *N. Engl. J. Med.* **372**, 2235–2242 (2015).
24. Zhang, J. *et al.* Germline mutations in predisposition genes in pediatric cancer. *N. Engl. J. Med.* **373**, 2336–2346 (2015).
25. Whirl-Carrillo, M. *et al.* Pharmacogenomics Knowledge for Personalized Medicine. *Clinical Pharmacology & Therapeutics* **92**, 414–417 (2012).
26. Kalia, S. S. *et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SFv2.0): a policy statement of the American college of medical genetics and genomics. *Genet. Med.* **19**, 249–255 (2016).
27. Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A. & McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* **33**, D514–517 (2005).
28. Mailman, M. D. *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **39**, 1181–1186 (2007).
29. Ledermann, J. *et al.* Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. *N. Engl. J. Med.* **366**, 1382–1392 (2012).
30. Yin, M. *et al.* ERCC1 and ERCC2 polymorphisms predict clinical outcomes of oxaliplatin-based chemotherapies in gastric and colorectal Cancer: a systemic review and meta-analysis. *Clin. Cancer Res.* **17**, 1632–1640 (2011).
31. Rosenthal, R. *et al.* “deconstructSigs: Delineating Mutational Processes in Single Tumors Distinguishes DNA Repair Deficiencies and Patterns of Carcinoma Evolution.” *Genome Biology* **17**, 31 (2016).
32. Dang, H. *et al.* ClonEvol: clonal ordering and visualization in cancer sequencing. *Annals of oncology* **28**, 3076–3082 (2017).
33. Mihalcea, D. J., Florescu, M. & Vinereanu, D. Mechanisms and genetic susceptibility of chemotherapy-induced cardiotoxicity in patients with breast cancer. *Am. J. Ther.* **24**, e3–11 (2017).
34. Aldubayan, S. H. *et al.* Enrichment of germline DNA-repair gene mutations in patients with colorectal cancer. *J. Clin. Oncol.* **35**, 1500 (2007).
35. Huang, K. L. *et al.* Pathogenic germline variants in 10,389 adult cancers. *Cell* **173**, 355–370 (2018).
36. Pearlman, R. *et al.* Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. *JAMA Oncol.* **3**, 464–471 (2017).
37. Elliott, A. M. Friedman JM3. The importance of genetic counselling in genome-wide sequencing. *Nat. Rev. Genet.* **19**, 735–736 (2018).
38. Vinson, K. E., George, D. C., Fender, A. W., Bertrand, F. E. & Sigounas, G. The Notch pathway in colorectal cancer. *Int. J. Cancer* **138**, 1835–1842 (2016).
39. Joung, J. G. *et al.* Tumor heterogeneity predicts metastatic potential in colorectal cancer. *Clin. Cancer Res.* **23**, 7209–7216 (2017).
40. Flanagan, S. E., Patch, A. M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet. Test Mol. Biomarkers* **14**, 533–537 (2010).
41. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* **17**, 405–24 (2015).

## Acknowledgements

The authors are grateful to the patients and their families at National Cheng Kung University Hospital. We also thank Kimforest LTD Taiwan for the bioinformatics support. This study was partly supported by the Ministry of Science and Technology, Ministry of Education, Ministry of Health and Welfare (MOHW106-TDU-B-211-113003), and National Cheng Kung University Hospital, Taiwan.

## Author Contributions

Peng-Chan Lin and Meng-Ru Shen designed study, supervised all phases of the project and wrote the manuscript. Yu-Min Yeh and Pei-Ying Wu provided coordination of clinical samples and testing and oversight of genetic counseling. Keng-Fu Hsu and Jang-Yang Chang provided interpretation of clinical analysis. All authors read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-40571-0>.

**Competing Interests:** The authors declare no competing interests.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019