

# 1 Comprehensive Study of Germline Mutations and Double-Hit Events 2 in Esophageal Squamous Cell Cancer

3 Bing Zeng<sup>1,2</sup>, Peide Huang<sup>2</sup>, Peina Du<sup>3</sup>, Xiaohui Sun<sup>3</sup>, Xuanlin Huang<sup>2</sup>, Xiaodong Fang<sup>1,2,4\*</sup>, Lin  
4 Li<sup>2\*</sup>

5 <sup>1</sup> BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, 518083, China

6 <sup>2</sup> BGI-Shenzhen, Shenzhen, 518083, China

7 <sup>3</sup> BGI Genomics, BGI-Shenzhen, Shenzhen, 518083, China

8 <sup>4</sup> China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

9 \* Correspondence:

10 Lin Li

11 Email: [lilin\\_contact@163.com](mailto:lilin_contact@163.com)

12 Xiaodong Fang

13 Email: [fangxd@bgi.com](mailto:fangxd@bgi.com)

14 **Keywords: Esophageal squamous cell cancer, cancer susceptibility gene, double-hit, germline**  
15 **mutation, pathogenicity**

16

## 17 Abstract

18 Esophageal squamous cell cancer (ESCC) is the eighth most common cancer around the world.  
19 Several reports have focused on somatic mutations and common germline mutations in ESCC.  
20 However, the contributions of pathogenic germline alterations in cancer susceptibility genes (CSGs),  
21 highly frequently mutated CSGs, and pathogenically mutated CSG-related pathways in ESCC remain  
22 unclear. We obtained data on 571 ESCC cases from public databases and East Asian from the 1000  
23 Genomes Project database and the China Metabolic Analytics Project database to characterize  
24 pathogenic mutations. We detected 157 mutations in 75 CSGs, accounting for 25.0% (143/571) of  
25 ESCC cases. Six genes had more than five mutations: *TP53* (n = 15 mutations), *GJB2* (n = 8),  
26 *BRCA2* (n = 6), *RECQL4* (n = 6), *MUTYH* (n = 6), and *PMS2* (n = 5). Our results identified  
27 significant differences in pathogenic germline mutations of *TP53*, *BRCA2*, and *RECQL4* between the  
28 ESCC and control cohorts. Moreover, we identified 84 double-hit events (16 germline/somatic  
29 double-hit events and 68 somatic/somatic double-hit events) occurring in 18 tumor suppressor genes  
30 from 83 patients. Patients who had ESCC with germline/somatic double-hit events were diagnosed at  
31 younger ages than patients with the somatic/somatic double-hit events, although the correlation was  
32 not significant. Fanconi anemia was the most enriched pathway of pathogenically mutated CSGs, and  
33 it appeared to be a primary pathway for ESCC predisposition. The results of this study identified the  
34 underlying roles that pathogenic germline mutations in CSGs play in ESCC pathogenesis; increased  
35 our awareness about the genetic basis of ESCC; and provided suggestions for using highly mutated  
36 CSGs and double-hit features in the early discovery, prevention, and genetic counseling of ESCC.

## 37 1 Introduction

38 Esophageal squamous cell cancer (ESCC) is one of the most common cancers in the world, and it is  
39 especially common in Asian countries, North America, and the eastern corridor of Africa (1). In  
40 China, there are approximately 478,000 new cases and 375,000 deaths related to ESCC each year (2).  
41 Many factors reportedly have relationships with ESCC; these include smoking, drinking, and dietary  
42 habits (3). However, the hereditary factors involved in ESCC remain unclear. Thus, understanding  
43 the genetic mutations and molecular events in ESCC might be pivotal to reduce the incidence and  
44 mortality rate of ESCC.

45 Enormous efforts have been taken to identify somatic alterations by whole-genome sequencing  
46 (WGS) or whole-exome sequencing (WES) (4,5), and several studies reveal the complex process of  
47 tumor development (6,7). Many common germline single-nucleotide polymorphisms (SNPs) have  
48 been identified by genome-wide association studies (8–16). rs138478634, a *CYP26B1* low-frequency  
49 variant, was proved to be involved in ESCC development (14). In 2018, several pan-cancer studies  
50 focused on pathogenic germline mutations to explore hereditary factors in cancers; 871 rare cancer  
51 predisposition mutations and copy number variations (CNVs) were observed in 8% of 10,389 cases,  
52 and 7.6% of the 914 patients with pediatric cancers had tumors that harbored pathogenic mutations in  
53 cancer predisposition genes (17,18). Deng et al. (19) identified germline profiles in Chinese patients  
54 with ESCC and uncovered the association between genotype and environment interactions.  
55 Additionally, *BRCA2* was associated with ESCC risk in Chinese patients (20). Reflecting a critical  
56 part of cancer susceptibility, the two-hit hypothesis assumes that hereditary retinoblastoma involves  
57 double mutations and that one mutation is in germline DNA whereas nonhereditary retinoblastoma  
58 involves two somatic mutations (21). On the basis of these findings, double-hit events in some  
59 studies were used to identify cancer predisposition genes (22,23). These studies demonstrated the  
60 significance of pathogenic germline mutations and double-hit events in genetic testing and risk  
61 assessment for cancer.

62 To our knowledge, cancer predisposition genes and molecular events in ESCC remain poorly  
63 understood. Here, we identified pathogenic/likely pathogenic germline predisposition mutations and  
64 highly frequently mutated CSGs in a large ESCC cohort. We discovered significantly different  
65 pathogenic germline mutations of *TP53*, *BRCA2*, and *RECQL4* in ESCC cohorts, and we clarified the  
66 association between double-hit events and diagnosis age in patients with ESCC. In addition, we  
67 identified pathogenically mutated CSG-related pathways for ESCC to illuminate the mechanism  
68 affected by pathogenic mutations. Results of this study will improve genetic testing for relatives of  
69 patients with ESCC and facilitate implementation of organizational or institutional measures for  
70 ESCC prevention and surveillance.

## 71 2 Materials and Methods

### 72 2.1 Sample acquisition

73 We collected 592 ESCC samples from published studies and The Cancer Genome Atlas (a total of  
74 nine projects) (Supplementary Table 1), and we excluded poor-quality samples and hypermutant  
75 samples (4,5,24–29). Clinical information is listed in Supplementary Table 2. The WGS and WES  
76 data from the same studies came from distinct patient cases. The quality control analysis uncovered  
77 an average sequencing depth of 55×~161× for WES samples and 30×~65× for WGS samples  
78 (Supplementary Figure 1A), the 10× average coverages were more than 90% in most WES and WGS  
79 samples (Supplementary Figure 1B). Moreover, the relationship between 10× average coverages and  
80 average sequencing depths showed a positive correlation (Supplementary Figure 1C), suggesting that

81 the qualities of most samples were proofed. The mean depth of our data and the public databases we  
82 used as controls were able to provide enough variants to execute downstream analysis (30). The  
83 study protocol was reviewed by the institutional review board of the Beijing Genomics Institution.

## 84 **2.2 Data processing and mutation calling**

85 The fastq data from 571 samples (38 WGS samples, 533 WES samples) were trimmed and filtered  
86 using SOAPnuke (v1.5.6 with default parameters, except where -n 0.1 -l 11 -q 0.5 -G -T 1) (31). Data  
87 from ESCC-P006 was transformed from bam files using the GATK SamToFastq (v4.0.6.0 with  
88 default parameters) (32). The high-quality reads were aligned to the hg19 human reference genome  
89 with a Burrows-Wheeler Aligner (v0.7.17-r1194-dirty with default parameters, except where -o 1 -e  
90 50 -m 100000 -i 15 -q 10 -a 600) (33). MarkDuplicates GATK (version as above with default  
91 parameters, except where -CREATE\_INDEX true, -reportMemoryStats true, -  
92 VALIDATION\_STRINGENCY SILENT) was used to mark duplicated reads. BaseRecalibrator  
93 (version as above with default parameters) and ApplyBQSR (version as above with default  
94 parameters, except where -create-output-bam-index true) were performed to base quality score  
95 recalibration (32). Germline variants were joint-called using GenotypeGVCFs (version as above with  
96 default parameters, except where -ignore-variants-starting-outside-interval true) after  
97 CombineGVCFs (version as above with default parameters) and annotated with the Variant Effect  
98 Predictor (VEP v98.3) (32,34). The calling germline variants of nine projects are shown in  
99 Supplementary Figure 1D. Samples with fewer than 80,000 variants were filtered out. Somatic  
100 variants were detected by GATK MuTect2 (version as above with default parameters except where -  
101 af-of-alleles-not-in-resource 0.0000025, -native-pair-hmm-threads 1, -add-output-vcf-command-line  
102 false), and Oncotator (v1.9.9.0) was used for annotation (32,35). Loss of heterozygosity (LOH) and  
103 other somatic CNVs (SCNVs) were detected with FACETS (v0.5.14) and Pathwork (v1.0) for 533  
104 WES and 38 WGS samples, respectively (36,37).

## 105 **2.3 CSG sets**

106 We curated CSGs from published papers and the Catalogue of Somatic Mutations in Cancer  
107 (COSMIC) database (38); we included cancer predisposition genes from three papers (17,18,39) and  
108 genes with recorded germline associations in COSMIC (Supplementary Table 4). After we removed  
109 duplicated genes, the CSG set included 260 genes. CSGs were divided into three groups according to  
110 the literature (17,40–42); these groups were tumor suppressor genes (TSGs; n = 139), oncogenes (n =  
111 36), and nonclassified genes (n = 85).

## 112 **2.4 Pathogenicity evaluation**

113 We first leveraged an in-house pathogenicity database to match germline variants; the rest of the  
114 germline variants were evaluated using InterVar (InterVar\_20190327) as a supplemental method to  
115 find germline pathogenic/likely pathogenic mutations (43). Germline pathogenic or likely pathogenic  
116 variants are hereafter referred to as pathogenic mutations. The pathogenicity database included  
117 ClinVar, the Human Gene Mutation Database, mutations collected from papers, and mutations we  
118 assessed according to consensus guidelines by the American College of Medical Genetics and  
119 Genomics and the Association for Molecular Pathology (17,44–46). We filtered for pathogenic  
120 variants with an allele frequency of 0.5% or lower in the Genome Aggregation Database (gnomAD  
121 version v2.1) (47). Pathogenic mutations in 260 high-interest CSGs (Supplementary Table 6) were  
122 selected for analysis and were checked by Deep Variant (48); manual verification ruled out false-  
123 positive results. For somatic nonsilent variants, with the exception of frameshift, nonsense, and  
124 splice-site mutations, three silico tools—SIFT (49), Polyphen2\_HDIV (50) and CADD (51) were

125 used to predict pathogenicity. If a variant was predicted as damaging in any two silico tools (SIFT:  
126 D, Polyphen2\_HDIV: D/P, CADD score > 15), the variant was categorized as deleterious (39,52).

## 127 **2.5 Identification of potential double-hit events**

128 According to the two-hit hypothesis, potential double-hit events are identified after two or more hits  
129 have been found in the same CSG; in this study, we set rigorous standards for determining hits.  
130 Pathogenic germline mutations were considered hits. Effective somatic variations were defined as  
131 hits if they met the following requirements: frameshift, nonsense, splice-site mutations, or deleterious  
132 missense and in-frame variants and SCNVs that caused allele loss. Copy-neutral LOH, duplication  
133 LOH, homozygous deletion, and hemizygous deletion were assumed to be linked to allele loss and  
134 were termed allele loss SCNVs (53,54). Integrative Genomics Viewer software was used to examine  
135 the authenticity of biallelic events (55). For double-hit events comprised of germline hits and allele  
136 loss SCNVs, we calculated SNP average depths and variant allelic frequency in normal and tumor  
137 tissues of ESCC to further validate allele loss SCNV events. Samples with variant allele frequencies  
138 less than 0.5 in tumors were removed.

## 139 **2.6 Statistical analyses**

140 To evaluate the correlations of the clinical features and genetic events, we used the two-sided  
141 Student's t-test. We conducted the two-sided Fisher's exact test to assess the gene-based association  
142 analysis and pathway enrichment. We also performed a burden test to determine the exact  
143 relationships between pathogenic mutations in CSGs and ESCC (56);  $p < 0.05$  was defined as  
144 statistically significant.

## 145 **3 Results**

### 146 **3.1 Population characteristics**

147 Overall, 469 of 571 patient cases were Asian (424 Chinese, 41 Vietnamese, one Canadian, one  
148 Brazilian, two without country information), 41 were Caucasian, 58 were Black or African American,  
149 and the rest were Brazilian without ethnicity information. The entire population consisted of 105  
150 women, 465 men, and one patient without gender information. The average diagnosed age for 567  
151 patients (the rest had no information) was 58.81 years (the minimum diagnosed age was 24 years,  
152 and the maximum diagnosed age was 93 years). Thirty-five patients had family histories of ESCC,  
153 and the average age of patients with ESCC with a family history (mean age and standard deviation  
154 [SD] was 56.80 [9.3] years; [range, 41–82 years]). This average was lower than the age of patients  
155 with ESCC without a family history (mean age [SD], 60.00 [8.2] years; [range, 36–78 years]; t-test  $p$   
156 = 0.059; 95% CI, -6.511 to 0.121) (Supplementary Figure 2). The average survival for 399 patients  
157 (the rest had no information) was 879.8 days (minimum survival, 3 days; maximum survival, 2,580  
158 days). In this study, 347 patients had a smoking history, and 215 patients had histories of alcoholism.  
159 With regard to disease grade, 334 patients had disease with pathological grade 2 or lower, and 86  
160 patients had disease with pathological grade greater than 2; pathological grade information was  
161 missing for 151 patients. All patients were diagnosed with disease stages I ( $n = 72$ ), stage II ( $n =$   
162 207), stage III ( $n = 203$ ), and stage IV ( $n = 7$ ); 82 patients were not assigned disease stages for this  
163 study (their information was lost).

164 [Insert Figure 1 here]

### 165 **3.2 Pathogenic germline mutations in CSGs**

166 Overall, 2,484 pathogenic germline mutations were identified, including 1,973 SNPs and 511  
167 insertions or deletions (Supplementary Table 5). Each sample had an average of 4.4 pathogenic  
168 mutations. After filtration by CSGs, 157 pathogenic mutations (113 SNPs and 44 insertions or  
169 deletions) were discovered from 25.0% (143/571) of the population (Supplementary Figure 3).  
170 Although each sample had an average of 1.1 pathogenic mutation in CSGs, only 12 (2.10%) of the  
171 571 patients harbored one or more pathogenic mutation in CSGs (Figure 1, Supplementary Table 6).  
172 The frequency of most mutations was rare in the gnomAD noncancer database and in the China  
173 Metabolic Analytics Project (ChinaMAP) database (47,57), indicating the sparsity of these  
174 deleterious mutations in the general population. As expected, most of the frequently mutated CSGs  
175 belonged to TSGs, and they were involved in biological processes, such as DNA repair.

176 In general, the CSGs detected more than five times were *TP53* (n = 15 mutations), *GJB2* (n = 8),  
177 *BRCA2* (n = 6), *RECQL4* (n = 6), *MUTYH* (n = 6), and *PMS2* (n = 5). *TP53* was the most frequently  
178 mutated CSG, with pathogenic germline mutations in 2.63% (15/571) of patients with ESCC (Figure  
179 1, Supplementary Table 6 and Supplementary Figure 4); The result was the same as *TP53* pathogenic  
180 mutations in a study of osteosarcoma (39). In our study, 86.7% (13/15) of *TP53* mutations were  
181 nonsynonymous single-nucleotide variations. c.A1073T (rs773553186; in 0.35%, or 2/571) and  
182 c.C742T (rs121912851; in 0.18%, or 1/571) were recorded in the International Agency for Research  
183 on Cancer TP53 database (58). All *TP53* pathogenic mutations were found in Chinese patients,  
184 except c.A1073T (one each in a Chinese and a Caucasian patient) (Supplementary Figure 4). Three of  
185 the *TP53* mutations, c.C742T, c.C586T, and c.C817T, have been reported in osteosarcoma (39), and  
186 *TP53* c.C742T has also been identified in low-grade glioma (17) (Supplementary Figure 4).  
187 Pathogenic mutations in *GJB2* represented the second most frequently mutated CSGs (Figure 1);  
188 their detection rate was 1.40% (8/571). The c.235delC (rs80338943) mutation, a common pathogenic  
189 frameshift deletion mutation in East Asian (EAS) populations, has been detected in six Asian  
190 (Chinese) patients with ESCC (59). Because this mutation has not been detected in other populations,  
191 rs80338943 may be specific to Chinese or Asian populations.

192 Nonsynonymous single-nucleotide variations occupied no less than 50% of pathogenic germline  
193 mutations in *BRCA2*, *RECQL4*, and *MUTYH* (Supplementary Table 6). In the upstream region, we  
194 detected a pathogenic splice mutation, *BRCA2* c.-39-1\_-39delGA (rs758732038), in a patient, and the  
195 mutation was reported in ClinVar as likely pathogenic (46). The mutation has also been reported in  
196 patients with breast cancer and medulloblastoma (60–62). *RECQL4* pathogenic mutations were only  
197 detected in Asian (Chinese) patients in our study, and *RECQL4* c.C2272T has been reported in  
198 ovarian cancer/Rothmund–Thomson syndrome. In our study, *MUTYH* c.C1178T (rs36053993) and  
199 c.C458T (rs762307622) were detected three times (0.53%, or 3/571) and two times (0.35%, or  
200 2/571), respectively. rs36053993 only detected in Caucasian patients and rs762307622 only detected  
201 in Asian (Chinese) patients. From gnomAD, rs36053993 in a homozygous state was found in three  
202 non-Finnish Europeans; this mutation may have been caused by founder events (63,64). Pathogenic  
203 mutations in *PMS2* were detected five times in five patients in our study (0.88%), and  
204 c.2192\_2196delAGTTA (rs63750695) was observed in only four patients, who were all African. The  
205 rs63750695 mutation has also been discovered in Lynch syndrome, colorectal cancer, and ovarian  
206 carcinoma (65–67); however, it was rare in noncancer gnomAD and ChinaMAP, for which  
207 frequencies were  $1.15 \times 10^{-5}$  and 0, respectively (Figure 1). rs63750695 is possibly specific to  
208 African ethnicity in ESCC.

209 The total number of pathogenic germline mutations and the frequency of mutations were relatively  
210 lower in oncogenes and nonclassified genes compared with TSGs. *TSHR* and *MPL* were oncogenes

211 that were mutated in two patients with ESCC; other oncogenes occurred in just one patient.  
212 *SLC25A13* was one of the nonclassified genes with the most pathogenic mutations.

213 We also investigated our pathogenic germline mutations in a previous pan-cancer study (17). Nine  
214 mutations were spread over 22 samples with diverse cancers (Supplementary Table 9). *SLC25A13*  
215 c.852\_855delCATA (n = 7), *GJB2* c.235delC (n = 7), and *PALB2* c.C2257T (n = 2) were the variants  
216 observed more than once across cancers. We detected multiple susceptibility loci (31/47), also  
217 identified in previous genome-wide association studies, in our patients with ESCC (Supplementary  
218 Table 10) (8–16). Of those genes with susceptibility loci, pathogenic mutations *PDE4D* c.T108A and  
219 *RUNX1* c.61+1delG were found in two patients separately (Supplementary Table 5). We also  
220 confirmed from the COSMIC database that 87.3% (137/157) of pathogenic mutations in CSGs had  
221 nonsilent somatic mutations in the same or a nearby (within five) amino acid position  
222 (Supplementary Table 6). Among 137 mutations, 107 mutations were observed in TSGs, representing  
223 89.2% (107/120) of all mutations.

224 [Insert Table 1 here]

### 225 3.3 Pathogenic germline mutations frequency in ESCC cases versus controls

226 To reveal the relationships between highly frequent mutated CSGs and ESCC, we chose the Chinese  
227 patients to continue the study, to leverage the most population data and avoid any ethnicity-specific  
228 effect. We conducted gene-based association analyses by comparing various germline mutation data  
229 from individuals with ESCC versus a 1000 Genomes Project EAS population and versus a  
230 ChinaMAP population separately (57,68). We also conducted rare variant burden tests on the ESCC  
231 individuals and the 1000 Genomes Project EAS population (68). Through the same pathogenicity  
232 evaluation pipeline, pathogenic mutations were identified in two public database populations.  
233 Analysis of results identified significantly higher pathogenic mutations in Chinese patients with  
234 ESCC versus public population databases (including 1000 Genomes Project EAS and ChinaMAP  
235 data), as reflected by odd ratios (ORs) of pathogenic mutations in *TP53* from the Chinese ESCC  
236 populations compared with the 1000 Genomes Project EAS populations (OR = 4.26; 95% CI, 1.33-  
237 17.91; Fisher's exact test  $p = 7.359 \times 10^{-3}$ ) and compared with the ChinaMAP populations (OR =  
238 10.59; 95% CI, 5.21-20.45; Fisher's exact test  $p = 1.851 \times 10^{-9}$ ); in *BRCA2* from the Chinese ESCC  
239 populations compared with the 1000 Genomes Project EAS populations (OR = infinity; 95% CI, 1.09-  
240 infinity; Fisher's exact test  $p = 0.0197$ ) and compared with the ChinaMAP populations (OR = 2.68;  
241 95% CI, 0.83-6.75; Fisher's exact test  $p = 0.0489$ ); and in *RECQL4* from the Chinese ESCC  
242 populations compared with the 1000 Genomes Project EAS populations (OR = 7.21; 95% CI, 0.87-  
243 332.231; Fisher's exact test  $p = 0.0519$ ) and compared with the ChinaMAP populations (OR = 3.69;  
244 95% CI, 1.27-8.81; Fisher's exact test  $p = 0.0089$ ) (Table 1). Likewise, in the burden analyses (Table  
245 1), the numbers of pathogenic mutations from *TP53* (14/424, or 3.30%; burden test  $p = 3.050 \times 10^{-3}$ ),  
246 *BRCA2* (5/424, or 1.18%; burden test  $p = 0.015$ ), and *RECQL4* (6/424, or 1.14%; burden test  $p =$   
247 0.035) in our Chinese ESCC cohort were higher than those observed in the 1000 Genomes Project  
248 EAS group.

249 [Insert Figure 2 here]

### 250 3.4 Potential double-hit events

251 To further survey the genetic predisposition of ESCC, we tried to identify potential double-hit events  
252 in ESCC. First, we identified 49,876 nonsilent mutations (Supplementary Table 3) in protein-coding  
253 regions from patients with ESCC. (We filtered the somatic mutations that overlapped with our own

254 panel of normal datasets and the Exome Aggregation Consortium database.) Then, by integrating  
255 pathogenic germline mutations and effective somatic mutations (Supplementary Table 8) or allele  
256 loss SCNVs, we found 84 potential double-hit events (Figure 2). To distinguish hits with germline  
257 mutations, the double-hit events were classified as germline/somatic double-hit events and  
258 somatic/somatic double-hit events. We identified 16 potential germline/somatic double-hit events  
259 (two germline mutations coupled with somatic mutations, and 14 germline mutations accompanied  
260 with allele loss SCNVs) (Figure 2, Supplementary Table 11 and Supplementary Figures 5, 6) in 16  
261 patients with ESCC, and we identified 68 potential somatic/somatic double-hit events (three somatic  
262 mutations accompanied by allele loss SCNVs and 65 double somatic mutations) (Figure 2,  
263 Supplementary Table 12) in 67 cases. The likelihood of two or more somatic mutations happening on  
264 the same chromosome was very low (52,69,70). Therefore, we assumed that double somatic  
265 mutations were likely in the trans position. Briefly, 83 individuals with ESCC possessed potential  
266 double-hit events, representing 14.5% of the ESCC cohort (Figure 2). Notably, one patient had two  
267 somatic/somatic double-hit events in different genes.

268 *GJB2* and *TP53* were the top two CSGs that found germline/somatic double-hit events.  
269 Germline/somatic double-hit events were identified in eight CSGs, including *BRCA2*, *BRCA1*,  
270 *MUTYH*, *CDKN2A*, and *ATM*. The dominant type of germline/somatic double-hit events was a  
271 germline mutation accompanied by an allele loss SCNv. In the remaining, germline mutations were  
272 coupled with somatic mutations; these were only discovered in *TP53* and *BRCA1*, possibly because  
273 SCNVs are relatively abundant in tumors and cover large genome region. In the somatic/somatic  
274 double-hit events, the *TP53* gene had the highest frequency, and most of the remaining genes had one  
275 potential double-hit event. Double somatic mutation was the main type of somatic/somatic double-hit  
276 event (Supplementary Table 12).

277 [Insert Figure 3 here]

278 When we compared diagnosis ages of patients with different double-hit events, we found that  
279 patients with germline/somatic double-hit events (with pathogenic germline mutations) had younger  
280 diagnosis ages (mean age [SD], 54.6 [11.2] years; [range, 36–71 years]) compared with patients in  
281 the somatic/somatic double-hit events (without pathogenic germline mutations; mean age [SD], 60.6  
282 [7.8] years; [range, 4–80 years]; t-test  $p = 0.056$ ; 95% CI, -12.216 to 0.177) (Figure 3). The  
283 comparison was nonsignificant, maybe it's due to the limited number of samples with double-hit  
284 events in this comparison. However, the finding was consistent in the study by Knudson (21). Using  
285 the empirical cumulative distribution function (ecdf) to calculate the expression percentiles of TSGs  
286 in an ESCC-P006 cancer cohort, two patients with somatic/somatic double-hit events showed low  
287 expression: one in *TP53* (5.32%) and one in *PTEN* (6.38%) (Supplementary Figure 8) (17). Those  
288 results support the two-hit hypothesis and suggest that genetic screening in specific TSGs can detect  
289 patients with germline/somatic double-hit events earlier.

290 [Insert Figure 4 here]

### 291 3.5 Pathway enrichment

292 To obtain a more comprehensive understanding of pathogenic germline genetic mutations affecting  
293 pathways, Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses were performed  
294 for multiple gene lists. The Fanconi anemia (FA) pathway was the most significantly enriched in the  
295 analysis of 75 pathogenically mutated CSGs (Fisher's exact test  $p = 6.634 \times 10^{-19}$ ) (Figure 4A,  
296 Supplementary Table 7). In addition, 1,226 pathogenic mutated genes and the genes involved in  
297 germline/somatic double-hit events were significantly enriched in this pathway. The top four

298 pathways for CSGs involved in somatic/somatic double-hit events versus for CSGs involved in  
299 germline/somatic double-hit events differed significantly (Supplementary Figures 7A-D).

300 In the tumor-suppressor network, the FA pathway functions to preserve genomic integrity by  
301 repairing DNA interstrand crosslinks, regulating cytokinesis, and mitigating replication stress  
302 (71,72). 33 ESCC samples carried pathogenic mutations in 13 CSGs included in the FA pathway  
303 (Figure 4B). The homologous recombination pathway and the mismatch repair pathway described in  
304 a previous ESCC project, and associated with cancer susceptibility, were found in our study  
305 (Supplementary Figure 7A) (19,73–75). Those pathways were also reported in pathway enrichments  
306 of ovarian cancer and osteosarcoma (39,76). We also interrogated the oncogenic signaling pathways  
307 upon which our mutated CSGs converged (77). The cell cycle pathway was the most enriched,  
308 followed by p53 pathway, the phosphatidylinositol 3'-kinase-Akt pathway, and the receptor tyrosine  
309 kinases-Ras pathway.

#### 310 **4 Discussion**

311 We reported the profile of pathogenic germline mutations of a larger ESCC cohort comparing with  
312 previous studies (17,19). We found 157 pathogenic mutations in CSGs from 143 (25.0%) of 571  
313 patients with ESCC and identified 84 double-hit events in 83 individuals (14.5%). Double-hit events  
314 were found in almost all projects in our study except ESCC-P008, which demonstrated that double-  
315 hit events are relatively common in ESCC. As far as we know, there was no report about pathogenic  
316 mutations in *GJB2*, *RECQL4*, *MUTYH* and *PMS2* in ESCC, however, they were discovered in our  
317 study. Overall, *TP53*, *GJB2*, *BRCA2*, *RECQL4*, *MUTYH*, and *PMS2* were highly frequently mutated  
318 CSGs. Significant pathways were identified for different CSGs with pathogenic mutations; the FA  
319 pathway appeared to be a primary pathway for cancer predisposition in ESCC. We showed that  
320 significantly more pathogenic mutations from *TP53*, *BRCA2* and *RECQL4* occurred in patients with  
321 ESCC than in control cohorts, which indicates that these three CSGs may play vital roles in ESCC.  
322 Interestingly, *TP53* and *RECQL4* have also been found significantly associated with osteosarcoma  
323 (39). The relationship with diagnosis age was not significant in our study, but double-hit events may  
324 be pivotal in ESCC carcinogenesis.

325 We found that *TP53* had the highest frequency of pathogenic germline mutations and the most  
326 double-hit events in CSGs. In our study, 80% (12/15) of germline mutations in *TP53* were located in  
327 the p53 domain, which functions in DNA binding. This domain contains four conserved regions that  
328 are enriched for somatic mutation hot spots and are essential for the function of the TP53 protein as a  
329 transcription factor (78,79). Six of the 12 mutations were discovered in conserved regions.  
330 Environmental factors and specific DNA sequences drive higher mutation rates, which may explain  
331 why p53 domain was a hot-spot region (80). Those pathogenic *TP53* mutations may disrupt the p53  
332 transcriptional pathway, which would enhance tumor progression and metastatic potential (81). The  
333 US Food and Drug Administration had approved drugs against the pocket in p53 domain (82). These  
334 drugs provide treatment options to patients with tumors that have mutations in the p53 domain.  
335 Results of studies in other cancers contrast with our findings about *TP53*. In a renal cell carcinoma  
336 study, *FH*, instead of *TP53*, harbored the most double-hit events, and *BRCA1* harbored the most in a  
337 pan-cancer study(17,22). Previous studies have reported that most double-hit events with *TP53*  
338 involve a mutation accompanied by LOH (83,84). However, in our research, double somatic  
339 mutations were the dominant type of double-hit event. It was partially due to the lack of researches  
340 on *TP53* double somatic mutations before.



341 *BRCA2* and *RECQL4* harbored more pathogenic germline mutations in ESCC than in public  
342 population. *BRCA2* is known for its involvement in breast cancer and ovarian cancer via the  
343 homologous recombination pathway, which is essential for repairing damaged DNA (85,86). And  
344 studies have reported *BRCA2* mutations related to ESCC risk in Chinese and Turkmen populations  
345 (20,87,88). The double-hit events detected in *BRCA2* in our study were germline/somatic double-hit  
346 events; the germline mutations were accompanied by allele loss SCNVs. These results were distinct  
347 from those reported in pancreatic acinar-cell carcinomas (89). *RECQL4* is a TSG that encodes  
348 RECQL4 helicase, which is involved in DNA replication and DNA repair. Germline mutations in  
349 *RECQL4* can cause Rothmund-Thomson syndrome and sporadic breast cancer (90). Although the  
350 pathogenic mutations in our ESCC cohort and in the 1000 Genomes EAS group were not  
351 significantly different (Fisher's exact test  $p = 0.0519$ ), the difference between them was also  
352 confirmed by analysis of the ChinaMAP cohort (Fisher's exact test  $p = 0.0089$ ). Importantly, this is  
353 the first report, to our knowledge, that illustrates the role of pathogenic mutations in *RECQL4* in  
354 ESCC.

355 The PMS2 protein is a homolog of the PMS1 protein (91) and both of them are components of the  
356 mismatch repair system. Common polymorphisms of *PMS1* have been positively associated with  
357 ESCC in an African population (92). This finding, together with the connection between PMS1 and  
358 PMS2, suggests a possible relationship between *PMS2* and ESCC. Double-hit events of mismatch  
359 repair genes could result in Lynch syndrome, as described in several studies (70,93), but we did not  
360 detect double-hit events in *PMS2* in our ESCC cohort. A larger ESCC cohort study might uncover  
361 double-hit events in *PMS2*, which would strengthen our understanding about ESCC susceptibility.

362 The genetic variations in ESCC are complicated. Although not all ESCC samples carried pathogenic  
363 germline mutations in CSGs, the detection rate of pathogenic mutations was close to that found in  
364 osteosarcoma (39). Because numerous susceptibility loci reported in genome-wide association  
365 studies were found in this research, we acknowledge that pathogenic mutations and known  
366 susceptibility loci may inform a genetic basis of ESCC. Our findings of variants and genes shared  
367 between ESCC and other cancers suggests that common hereditary factors exist in pan-cancer. Given  
368 the interplay of common SNPs and pathogenic mutations reported in breast cancer and colorectal  
369 cancer, the interaction between susceptibility loci and pathogenic mutations in ESCC suggests a need  
370 for future exploration (94).

371 To better understand the genetic factors causing ESCC initiation and development, we confirmed the  
372 putative germline-somatic interplay by COSMIC proximity match. The results not only support the  
373 pathogenicity of those germline mutations but also imply a signal functional relevance between  
374 germline and somatic mutations (76). In addition, we identified potential double-hit events in 83  
375 patients with ESCC; though the difference was not significant, the patients with germline/somatic  
376 double-hit events were more likely to be diagnosed at younger ages. It is possible that pathogenic  
377 mutations confer the earliest genetic hits to TSGs in cells, so a somatic hit alone would cause loss of  
378 function in TSGs (95). As a result of double-hit events, the cells generate malignancy. Furthermore,  
379 enriched pathways revealed the process of pathogenic mutations that affect ESCC tumorigenesis and  
380 development. In patients without pathogenic mutations or double-hit events, limited CSG sets,  
381 potential alternations in methylations of a promoter region, germline CNVs, and gene-environmental  
382 or gene-lifestyle interactions are possible explanations for ESCC development.

383 Despite our findings about the genetic characterization of and double-hit events in ESCC, we still  
384 acknowledge limitations to our study. The first is our inability to obtain detailed clinical information  
385 because of limited access to public databases. Second, merging different data, such as WGS and

386 WES, may induce biases in cohort-wide variant processing. Third, directly adopting variants from  
387 different sources may influence comparisons, because the different sources applied distinct platforms  
388 and variant detection pipelines. Fourth, our sample size was not large enough for statistical tests,  
389 especially for individual variants.

390 In sum, we report that approximately 25.0% of patients with ESCC harbored at least one pathogenic  
391 germline mutation in CSGs, and approximately 14.5% of ESCC cases could be explained by a two-  
392 hit hypothesis. Significantly enriched pathways also validated the significance of those pathogenic  
393 mutations. Myriad genome variations occur in patients; our findings represent, to our knowledge, the  
394 largest discovery of rare, germline predisposition mutations in ESCC so far. These results strengthen  
395 the understanding about genetic factors involved in ESCC and will help improve prevention, early  
396 detection, and risk management of ESCC for patients. We acknowledge the shortcomings in the  
397 analytical methods and the data sources used. Additional studies are needed to improve our  
398 observations and results.

## 399 **5 Data availability statement**

400 The raw data of this project can be found in the Sequence Read Archive hosted by the National  
401 Center for Biotechnology Information under accession numbers PRJNA315775, PRJNA399748,  
402 PRJNA401209, PRJNA230271, PRJNA317404, SRA112617 and SRP034680 and in the European  
403 Genome-Phenome Archive under accession number EGAD00001000845. The whole-exome  
404 sequencing data of esophageal squamous cell cancer samples from The Cancer Genome Atlas are  
405 available from the National Cancer Institute Genomic Data Commons  
406 (<https://portal.gdc.cancer.gov/>). All relevant datasets for this study are available from the authors.

## 407 **6 Conflict of Interest**

408 *The authors declare that the research was conducted in the absence of any commercial or financial*  
409 *relationships that could be construed as a potential conflict of interest.*

## 410 **7 Author Contributions**

411 LL and BZ: conceptualization. BZ: writing manuscript and performing analysis. PD, XS and XH:  
412 providing help in analysis. BZ, LL, XH and PD: collected the data from published literature or  
413 database. P H revised the manuscript. LL and XF supervised and supported this project.

## 414 **8 Funding**

415 No funding.

## 416 **9 Acknowledgments**

417 This study makes use of data generated by the Molecular Oncology Laboratory of Prof. Qimin Zhan,  
418 the Translational Medicine Research Center, Shanxi Medical University of Prof. Yongping Cui, the  
419 Department of Radiation Oncology, Fudan University Shanghai Cancer Center of Prof. Kuaile Zhao,  
420 the Department of Bioinformatics and Computational Biology, The University of Texas MD  
421 Anderson Cancer Center of Prof. Han Liang, The Lineberger Comprehensive Cancer Center,  
422 University of North Carolina School of Medicine, Prof. Norman E. Sharpless, the Institute of Clinical  
423 Pathology, Shantou University Medical College, Prof. Min Su, the Cedars-Sinai Medical Center,  
424 UCLA School of Medicine, Prof. H. Phillip Koeffler and Prof. Jie He of Cancer Institute and

425 Hospital Chinese Academy of Medical Sciences. We also acknowledge other Professors for sharing  
426 the fastq data, The National Center for Biotechnology Information, The European Genome-phenome  
427 Archive, The Cancer Genome Atlas for sharing the esophageal squamous cell cancer data.

428 The last but not the least, I want to thanks to my wife Panhong Liu, without her support, without my  
429 scientific research.

## 430 **10 References**

- 431 1. Brown J, Stepien AJ, Willem P. Landscape of copy number aberrations in esophageal  
432 squamous cell carcinoma from a high endemic region of South Africa. *BMC cancer* (2020)  
433 **20**:281. doi:10.1186/s12885-020-06788-3
- 434 2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer  
435 statistics in China, 2015. *CA: A Cancer Journal for Clinicians* (2016) **66**:115–132.  
436 doi:10.3322/caac.21338
- 437 3. Engel LS, Chow WH, Vaughan TL, Gammon MD, Risch HA, Stanford JL, Schoenberg JB,  
438 Mayne ST, Dubrow R, Rotterdam H, et al. Population attributable risks of esophageal and  
439 gastric cancers. *Journal of the National Cancer Institute* (2003) **95**:1404–1413.  
440 doi:10.1093/jnci/djg047
- 441 4. Song Y, Li L, Ou Y, Gao Z, Li E, Li X, Zhang W, Wang JJ, Xu L, Zhou Y, et al.  
442 Identification of genomic alterations in oesophageal squamous cell cancer. *Nature* (2014)  
443 **508**:91–95. doi:10.1038/nature13176
- 444 5. Gao YB, Chen ZL, Li JG, Hu X da, Shi XJ, Sun ZM, Zhang F, Zhao ZR, Li ZT, Liu ZY, et  
445 al. Genetic landscape of esophageal squamous cell carcinoma. *Nature Genetics* (2014)  
446 **46**:1097–1102. doi:10.1038/ng.3076
- 447 6. Chen XX, Zhong Q, Liu Y, Yan SM, Chen ZH, Jin SZ, Xia TL, Li RY, Zhou AJ, Su Z, et al.  
448 Genomic comparison of esophageal squamous cell carcinoma and its precursor lesions by  
449 multi-region whole-exome sequencing. *Nature Communications* (2017) **8**:  
450 doi:10.1038/s41467-017-00650-0
- 451 7. Liu X, Zhang M, Ying S, Zhang C, Lin R, Zheng J, Zhang G, Tian D, Guo Y, Du C, et al.  
452 Genetic Alterations in Esophageal Tissues From Squamous Dysplasia to Carcinoma.  
453 *Gastroenterology* (2017) **153**:166–177. doi:10.1053/j.gastro.2017.03.033
- 454 8. Cui R, Kamatani Y, Takahashi A, Usami M, Hosono N, Kawaguchi T, Tsunoda T, Kamatani  
455 N, Kubo M, Nakamura Y, et al. Functional Variants in ADH1B and ALDH2 Coupled With  
456 Alcohol and Smoking Synergistically Enhance Esophageal Cancer Risk. *Gastroenterology*  
457 (2009) **137**:1768–1775. doi:10.1053/j.gastro.2009.07.070
- 458 9. Wang LD, Zhou FY, Li XMXCM, Sun LD, Song X, Jin Y, Li JLM, Kong GQ, Qi H, Cui J,  
459 et al. Genome-wide association study of esophageal squamous cell carcinoma in chinese  
460 subjects identifies a susceptibility locus at PLCE1. *Nature Genetics* (2010) **42**:759–765.  
461 doi:10.1038/ng.648

- 462 10. Wu C, Hu Z, He Z, Jia W, Wang F, Zhou Y, Liu Z, Zhan Q, Liu Y, Yu D, et al. Genome-  
463 wide association study identifies three new susceptibility loci for esophageal squamous-cell  
464 carcinoma in Chinese populations. *Nature Genetics* (2011) **43**:679–684. doi:10.1038/ng.849
- 465 11. Wu C, Kraft P, Zhai K, Chang J, Wang Z, Li Y, Hu Z, He Z, Jia W, Abnet CC, et al.  
466 Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify  
467 multiple susceptibility loci and gene-environment interactions. *Nature Genetics* (2012)  
468 **44**:1090–1097. doi:10.1038/ng.2411
- 469 12. Wu C, Wang Z, Song X, Feng XS, Abnet CC, He J, Hu N, Zuo XB, Tan W, Zhan Q, et al.  
470 Joint analysis of three genome-wide association studies of esophageal squamous cell  
471 carcinoma in Chinese populations. *Nature Genetics* (2014) **46**:1001–1006.  
472 doi:10.1038/ng.3064
- 473 13. Lin D, Wu C, Li D, Jia W, Hu Z, Zhou Y, Yu D, Tong T, Wang M, Lin D, et al. Genome-  
474 wide association study identifies common variants in SLC39A6 associated with length of  
475 survival in esophageal squamous-cell carcinoma. *Nature Genetics* (2013) **45**:632–638.  
476 doi:10.1038/ng.2638
- 477 14. Chang J, Zhong R, Tian J, Li J, Zhai K, Ke J, Lou J, Chen W, Zhu B, Shen N, et al. Exome-  
478 wide analyses identify low-frequency variant in CYP26B1 and additional coding variants  
479 associated with esophageal squamous cell carcinoma. *Nature Genetics* (2018) **50**:338–343.  
480 doi:10.1038/s41588-018-0045-8
- 481 15. Hu JL, Hu XL, Lu CX, Chen XJ, Fu L, Han Q, Cang SD. Variants in the 3'-untranslated  
482 region of CUL3 is associated with risk of esophageal squamous cell carcinoma. *Journal of*  
483 *Cancer* (2018) **9**:3647–3650. doi:10.7150/jca.27052
- 484 16. Suo C, Yang Y, Yuan Z, Zhang T, Yang X, Qing T, Gao P, Shi L, Fan M, Cheng H, et al.  
485 Alcohol Intake Interacts with Functional Genetic Polymorphisms of Aldehyde Dehydrogenase  
486 (ALDH2) and Alcohol Dehydrogenase (ADH) to Increase Esophageal Squamous Cell Cancer  
487 Risk. *Journal of Thoracic Oncology* (2019) **14**:712–725. doi:10.1016/j.jtho.2018.12.023
- 488 17. Huang K lin, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, Paczkowska M, Reynolds S,  
489 Wyczalkowski MA, Oak N, et al. Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell*  
490 (2018) **173**:355-370.e14. doi:10.1016/j.cell.2018.03.039
- 491 18. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, Johann  
492 PD, Balasubramanian GP, Segura-Wang M, Brabetz S, et al. The landscape of genomic  
493 alterations across childhood cancers. *Nature* (2018) **555**:321–327. doi:10.1038/nature25480
- 494 19. Deng J, Weng X, Ye J, Zhou D, Liu Y, Zhao K. Identification of the germline mutation  
495 profile in esophageal squamous cell carcinoma by whole exome sequencing. *Frontiers in*  
496 *Genetics* (2019) **10**:1–10. doi:10.3389/fgene.2019.00047
- 497 20. Ko JMY, Ning L, Zhao XK, Chai AWY, Lei LC, Choi SSA, Tao L, Law S, Kwong A, Lee  
498 NPY, et al. BRCA2 loss-of-function germline mutations are associated with esophageal  
499 squamous cell carcinoma risk in Chinese. *International Journal of Cancer* (2020) **146**:1042–  
500 1051. doi:10.1002/ijc.32619

- 501 21. Knudson AG. Two genetic hits (more or less) to cancer. *Nature Reviews Cancer* (2001)  
502 1:157–162. doi:10.1038/35101031
- 503 22. Kemel Y, Zhang L, Knezevic A, Patil S, Ceyhan-birsoy O, Huang K, Redzematovic A,  
504 Coskey DT, Stewart C, Pradhan N, et al. Prevalence of Germline Mutations in Cancer  
505 Susceptibility Genes in Patients With Advanced Renal Cell Carcinoma. (2018) **10065**:  
506 doi:10.1001/jamaoncol.2018.1986
- 507 23. Park S, Supek F, Lehner B. Systematic discovery of germline cancer predisposition genes  
508 through the identification of somatic second hits. *Nature Communications* (2018) **9**:  
509 doi:10.1038/s41467-018-04900-7
- 510 24. Lin DC, Hao JJ, Nagata Y, Xu L, Shang L, Meng X, Sato Y, Okuno Y, Varela AM, Ding  
511 LW, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma.  
512 *Nature Genetics* (2014) **46**:467–473. doi:10.1038/ng.2935
- 513 25. Zhang L, Zhou Y, Cheng C, Cui H, Cheng L, Kong P, Wang JJJJ, Li YY, Chen W, Song B,  
514 et al. Genomic analyses reveal mutational signatures and frequently altered genes in  
515 esophageal squamous cell carcinoma. *American Journal of Human Genetics* (2015) **96**:597–  
516 611. doi:10.1016/j.ajhg.2015.02.017
- 517 26. Hao JJ, Lin DC, Dinh HQ, Mayakonda A, Jiang YY, Chang C, Jiang Y, Lu CC, Shi ZZ, Xu  
518 X, et al. Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal  
519 squamous cell carcinoma. *Nature Genetics* (2016) **48**:1500–1507. doi:10.1038/ng.3683
- 520 27. Liu W, Snell JM, Jeck WR, Hoadley KA, Wilkerson MD, Parker JS, Patel N, Mlombe YB,  
521 Mulima G, Liomba NG, et al. Subtyping sub-Saharan esophageal squamous cell carcinoma by  
522 comprehensive molecular analysis. *JCI Insight* (2016) **1**:1–11. doi:10.1172/jci.insight.88755
- 523 28. Deng J, Chen H, Zhou D, Zhang J, Chen Y, Liu Q, Ai D, Zhu H, Chu L, Ren W, et al.  
524 Comparative genomic analysis of esophageal squamous cell carcinoma between Asian and  
525 Caucasian patient populations. *Nature Communications* (2017) **8**:1–9. doi:10.1038/s41467-  
526 017-01730-x
- 527 29. Kim J, Bowlby R, Mungall AJ, Robertson AG, Odze RD, Cherniack AD, Shih J, Pedamallu  
528 CS, Cibulskis C, Dunford A, et al. Integrated genomic characterization of oesophageal  
529 carcinoma. *Nature* (2017) **541**:169–174. doi:10.1038/nature20805
- 530 30. Ajay SS, Parker SCJ, Abaan HO, Fuentes Fajardo K v., Margulies EH. Accurate and  
531 comprehensive sequencing of personal genomes. *Genome Research* (2011) **21**:1498–1505.  
532 doi:10.1101/gr.123638.111
- 533 31. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y. SOAPnuke : A MapReduce Acceleration  
534 supported Software for integrated Quality Control and Preprocessing of High-Throughput  
535 Sequencing Data. *GigaScience* (2018) **7**:gix120. doi:10.1093/gigascience/gix120/4689118
- 536 32. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K,  
537 Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: A MapReduce framework  
538 for analyzing next-generation DNA sequencing data. *Genome Research* (2010) **20**:1297–1303.  
539 doi:10.1101/gr.107524.110

- 540 33. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform.  
541 *Bioinformatics* (2010) **26**:589–595. doi:10.1093/bioinformatics/btp698
- 542 34. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F.  
543 The Ensembl Variant Effect Predictor. *Genome Biology* (2016) **17**:1–14. doi:10.1186/s13059-  
544 016-0974-4
- 545 35. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ. Oncotator : Cancer Variant  
546 Annotation Tool. *Human Mutation* (2015) **36**:E2423–E2429. doi:10.1002/humu.22771
- 547 36. Mayrhofer M, DiLorenzo S, Isaksson A. Patchwork: Allele-specific copy number analysis of  
548 whole-genome sequenced tumor tissue. *Genome Biology* (2013) **14**:R24. doi:10.1186/gb-  
549 2013-14-3-r24
- 550 37. Shen R, Seshan VE. FACETS: Allele-specific copy number and clonal heterogeneity  
551 analysis tool for high-throughput DNA sequencing. *Nucleic Acids Research* (2016) **44**:1–9.  
552 doi:10.1093/nar/gkw520
- 553 38. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson  
554 E, Ponting L, et al. COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids  
555 Research* (2017) **45**:D777–D783. doi:10.1093/nar/gkw1121
- 556 39. Mirabello L, Zhu B, Koster R, Karlins E, Dean M, Yeager M, Gianferante M, Spector LG,  
557 Morton LM, Karyadi D, et al. Frequency of pathogenic germline variants in cancer-  
558 susceptibility genes in patients with osteosarcoma. *JAMA Oncology* (2020) **6**:724–734.  
559 doi:10.1001/jamaoncol.2020.0197
- 560 40. Zhao M, Sun J, Zhao Z. TSGene: A web resource for tumor suppressor genes. *Nucleic Acids  
561 Research* (2013) **41**:970–976. doi:10.1093/nar/gks937
- 562 41. Zhao M, Kim P, Mitra R, Zhao J, Zhao Z. TSGene 2.0: An updated literature-based  
563 knowledgebase for Tumor Suppressor Genes. *Nucleic Acids Research* (2016) **44**:D1023–  
564 D1031. doi:10.1093/nar/gkv1268
- 565 42. Liu Y, Sun J, Zhao M. ONGene: A literature-based database for human oncogenes. *Journal  
566 of Genetics and Genomics* (2017) **44**:119–121. doi:10.1016/j.jgg.2016.12.004
- 567 43. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-  
568 AMP Guidelines. *American Journal of Human Genetics* (2017) **100**:267–280.  
569 doi:10.1016/j.ajhg.2017.01.004
- 570 44. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E,  
571 Spector E, et al. Standards and guidelines for the interpretation of sequence variants: a joint  
572 consensus recommendation of the American College of Medical Genetics and Genomics and  
573 the Association for Molecular Pathology. *Genetics in Medicine* (2015) **17**:405–423.  
574 doi:10.1038/gim.2015.30
- 575 45. Stenson PD, Mort M, Ball E v., Evans K, Hayden M, Heywood S, Hussain M, Phillips AD,  
576 Cooper DN. The Human Gene Mutation Database: towards a comprehensive repository of

- 577 inherited mutation data for medical research, genetic diagnosis and next-generation sequencing  
578 studies. *Human Genetics* (2017) **136**:665–677. doi:10.1007/s00439-017-1779-6
- 579 46. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman  
580 D, Jang W, et al. ClinVar: Improving access to variant interpretations and supporting  
581 evidence. *Nucleic Acids Research* (2018) **46**:D1062–D1067. doi:10.1093/nar/gkx1153
- 582 47. Lek M, Karczewski KJ, Minikel E v., Samocha KE, Banks E, Fennell T, O’Donnell-Luria  
583 AH, Ware JS, Hill AJ, Cummings BB, et al. Analysis of protein-coding genetic variation in  
584 60,706 humans. *Nature* (2016) **536**:285–291. doi:10.1038/nature19057
- 585 48. Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J,  
586 Nguyen N, Afshar PT, et al. A universal snp and small-indel variant caller using deep neural  
587 networks. *Nature Biotechnology* (2018) **36**:983. doi:10.1038/nbt.4235
- 588 49. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on  
589 protein function using the SIFT algorithm. *Nature Protocols* (2009) **4**:1073–1082.  
590 doi:10.1038/nprot.2009.86
- 591 50. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS,  
592 Sunyaev SR. A method and server for predicting damaging missense mutations. *Nature*  
593 *Methods* (2010) **7**:248–249. doi:10.1038/nmeth0410-248
- 594 51. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: Predicting the  
595 deleteriousness of variants throughout the human genome. *Nucleic Acids Research* (2019)  
596 **47**:D886–D894. doi:10.1093/nar/gky1016
- 597 52. Geurts-Giele WRR, Leenen CHM, Dubbink HJ, Meijssen IC, Post E, Sleddens HFBM,  
598 Kuipers EJ, Goverde A, van den Ouweland AMW, van Lier MGF, et al. Somatic aberrations  
599 of mismatch repair genes as a cause of microsatellite-unstable cancers. *Journal of Pathology*  
600 (2014) **234**:548–559. doi:10.1002/path.4419
- 601 53. Cox C, Bignell G, Greenman C, Stabenau A, Warren W, Stephens P, Davies H, Watt S,  
602 Teague J, Edkins S, et al. A survey of homozygous deletions in human cancer genomes.  
603 *Proceedings of the National Academy of Sciences of the United States of America* (2005)  
604 **102**:4542–4547. doi:10.1073/pnas.0408593102
- 605 54. Ryland GL, Doyle MA, Goode D, Boyle SE, Choong DYH, Rowley SM, Li J, Bowtell DD,  
606 Tothill RW, Campbell IG, et al. Loss of heterozygosity: What is it good for? *BMC Medical*  
607 *Genomics* (2015) **8**:1–12. doi:10.1186/s12920-015-0123-z
- 608 55. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the  
609 integrative genomics viewer. *Cancer Research* (2017) **77**:e31–e34. doi:10.1158/0008-  
610 5472.CAN-17-0337
- 611 56. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for  
612 the combined effect of rare and common variants. *American Journal of Human Genetics*  
613 (2013) **92**:841–853. doi:10.1016/j.ajhg.2013.04.015

- 614 57. Cao Y, Li L, Xu M, Feng Z, Sun X, Lu J, Xu Y, Du P, Wang T, Hu R, et al. The ChinaMAP  
615 analytics of deep whole genome sequences in 10,588 individuals. *Cell Research* (2020)  
616 doi:10.1038/s41422-020-0322-9
- 617 58. Olivier M, Eeles R, Hollstein M, Khan MA, Harris CC, Hainaut P. The IARC TP53  
618 database: New online mutation analysis and recommendations to users. *Human Mutation*  
619 (2002) **19**:607–614. doi:10.1002/humu.10081
- 620 59. Dzhemileva LU, Barashkov NA, Posukh OL, Khusainova RI, Akhmetova VL, Kutuev IA,  
621 Gilyazova IR, Tadinova VN, Fedorova SA, Khidiyatova IM, et al. Carrier frequency of GJB2  
622 gene mutations c.35delG, c.235delC and c.167delT among the populations of Eurasia. *Journal*  
623 *of Human Genetics* (2010) **55**:749–754. doi:10.1038/jhg.2010.101
- 624 60. Kwong A, Shin VY, Ho JCW, Kang E, Nakamura S, Teo SH, Lee ASG, Sng JH, Ginsburg  
625 OM, Kurian AW, et al. Comprehensive spectrum of BRCA1 and BRCA2 deleterious  
626 mutations in breast cancer in Asian countries. *Journal of Medical Genetics* (2016) **53**:15–23.  
627 doi:10.1136/jmedgenet-2015-103132
- 628 61. Waszak SM, Northcott PA, Buchhalter I, Robinson GW, Sutter C, Groebner S, Grund KB,  
629 Brugières L, Jones DTW, Pajtler KW, et al. Spectrum and prevalence of genetic predisposition  
630 in medulloblastoma: a retrospective genetic study and prospective validation in a clinical trial  
631 cohort. *The Lancet Oncology* (2018) **19**:785–798. doi:10.1016/S1470-2045(18)30242-0
- 632 62. Wen WX, Allen J, Lai KN, Mariapun S, Hasan SN, Ng PS, Lee DSC, Lee SY, Yoon SY,  
633 Lim J, et al. Inherited mutations in BRCA1 and BRCA2 in an unselected multiethnic cohort of  
634 Asian patients with breast cancer and healthy controls from Malaysia. *Journal of Medical*  
635 *Genetics* (2018) **55**:97–103. doi:10.1136/jmedgenet-2017-104947
- 636 63. Aretz S, Tricarico R, Papi L, Spier I, Pin E, Horpaopan S, Cordisco EL, Pedroni M, Stienen  
637 D, Gentile A, et al. MUTYH-associated polyposis (MAP): Evidence for the origin of the  
638 common European mutations p.Tyr179Cys and p.Gly396Asp by founder events. *European*  
639 *Journal of Human Genetics* (2014) **22**:923–929. doi:10.1038/ejhg.2012.309
- 640 64. Taki K, Sato Y, Nomura S, Ashihara Y, Kita M, Tajima I, Sugano K, Arai M. Mutation  
641 analysis of MUTYH in Japanese colorectal adenomatous polyposis patients. *Familial Cancer*  
642 (2016) **15**:261–265. doi:10.1007/s10689-015-9857-1
- 643 65. Klift HM van der, Tops ĀCMJ, Bik EC, Boogaard MW, Borgstein A, Hansson KBM,  
644 Ausems MGEM, Garcia EG, Green A, Hes FJ, et al. Quantification of Sequence Exchange  
645 Events between PMS2 and PMS2CL Provides a Basis for Improved Mutation Scanning of  
646 Lynch Syndrome Patients. *Human Mutation* (2010) **31**:578–587. doi:10.1002/humu.21229
- 647 66. Zhang P, Kitchen-Smith I, Xiong L, Stracquadanio G, Brown K, Richter P, Wallace M, Bond  
648 E, Sahgal N, Moore S, et al. Germline and somatic genetic variants in the p53 pathway interact  
649 to affect cancer risk, progression and drug response. *bioRxiv* (2019)835918.  
650 doi:10.1101/835918
- 651 67. Staninova-Stojovska M, Matevska-Geskovska N, Panovski M, Angelovska B, Mitrevski N,  
652 Risteovski M, Jovanovic R, Dimovski A. Molecular Basis of Inherited Colorectal Carcinomas



- 653 in the Macedonian Population: An Update. *Balkan J Med Genet* (2019) **22**:5–16.  
654 doi:10.2478/bjmg-2019-0025
- 655 68. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, Clark AG,  
656 Donnelly P, Eichler EE, Flicek P, et al. A global reference for human genetic variation. *Nature*  
657 (2015) **526**:68–74. doi:10.1038/nature15393
- 658 69. Boland CR, Goel A. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* (2010)  
659 **138**:2073–2087.e3. doi:10.1053/j.gastro.2009.12.064
- 660 70. Sourrouille I, Coulet F, Lefevre JH, Colas C, Eyries M, Svrcek M, Bardier-Dupas A, Parc Y,  
661 Soubrier F. Somatic mosaicism and double somatic hits can lead to MSI colorectal tumors.  
662 *Familial Cancer* (2013) **12**:27–33. doi:10.1007/s10689-012-9568-9
- 663 71. Ceccaldi R, Sarangi P, D’Andrea AD. The Fanconi anaemia pathway: new players and new  
664 functions. *Nature reviews Molecular cell biology* (2016) **17**:337. doi:10.1038/nrm.2016.48
- 665 72. Joshi Niraj, Anniina Färkkilä, D’Andrea AD. The Fanconi Anemia Pathway in Cancer. *Annu*  
666 *Rev Cancer Biol* (2019) **3**:457–478. doi:10.1016/j.physbeh.2017.03.040
- 667 73. Hsieh P, Yamane K. DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mech*  
668 *Ageing Dev* (2008) **129**:391–407. doi:10.1016/j.mad.2008.02.012
- 669 74. Li GM. Mechanisms and functions of DNA mismatch repair. *Cell Research* (2008) **18**:85–  
670 98. doi:10.1038/cr.2007.115
- 671 75. Li X, Heyer W-D. Homologous recombination in DNA repair and DNA tolerance. *Cell*  
672 *Research* (2008) **18**:99–113. doi:10.1038/cr.2008.1
- 673 76. Kanchi KL, Johnson KJ, Lu C, McLellan MD, Leiserson MDM, Wendl MC, Zhang Q,  
674 Koboldt DC, Xie M, Kandoth C, et al. Integrated analysis of germline and somatic variants in  
675 ovarian cancer. *Nature Communications* (2014) **5**:1–14. doi:10.1038/ncomms4156
- 676 77. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL,  
677 Kantheti HS, Saghafeinia S, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas.  
678 *Cell* (2018) **173**:321–337.e10. doi:10.1016/j.cell.2018.03.035
- 679 78. Pavletich NP, Chambers KA, Pabo CO. The DNA-binding domain of 53 contains the four  
680 conserved regions the major mutation hot spots. *Genes & development* (1993) **7**:2556–2564.
- 681 79. Harms KL, Chen X. The functional domains in p53 family proteins exhibit both common and  
682 distinct properties. *Cell Death and Differentiation* (2006) **13**:890–897.  
683 doi:10.1038/sj.cdd.4401904
- 684 80. Baugh EH, Ke H, Levine AJ, Bonneau RA, Chan CS. Why are there hotspot mutations in the  
685 TP53 gene in human cancers? *Cell Death and Differentiation* (2018) **25**:154–160.  
686 doi:10.1038/cdd.2017.180
- 687 81. Parrales A, Iwakuma T. Targeting oncogenic mutant p53 for cancer therapy. *Frontiers in*  
688 *Oncology* (2015) **5**:1–13. doi:10.3389/fonc.2015.00288

- 689 82. Pradhan MR, Siau JW, Kannan S, Nguyen MN, Ouaray Z, Kwoh CK, Lane DP, Ghadessy F,  
690 Verma S. Simulations of mutant p53 DNA binding domains reveal a novel druggable pocket.  
691 *Nucleic Acids Research* (2019) **47**:1637–1652. doi:10.1093/nar/gky1314
- 692 83. Kurose K, Gilley K, Matsumoto S, Watson PH, Zhou XP, Eng C. Frequent somatic  
693 mutations in PTEN and TP53 are mutually exclusive in the stroma of breast carcinomas.  
694 *Nature Genetics* (2002) **32**:355–357. doi:10.1038/ng1013
- 695 84. Liu Y, Chen C, Xu Z, Scuoppo C, Rillaan CD, Gao J, Spitzer B, Bosbach B, Kastenhuber  
696 ER, Baslan T, et al. Deletions linked to TP53 loss drive cancer through p53-independent  
697 mechanisms. *Nature* (2016) **531**:471–475. doi:10.1038/nature17157
- 698 85. Buisson R, Dion-Côté A-M, Coulombe Y, Launay H. Cooperation of breast cancer proteins  
699 PALB2 and piccolo BRAC2 in stimulating homologous recombination. *Nature structural &  
700 molecular biology* (2010) **17**:1247–1254. doi:10.1038/nsmb.1915
- 701 86. Girardi F, Barnes DR, Barrowdale D, Frost D, Brady AF, Miller C, Henderson A, Donaldson  
702 A, Murray A, Brewer C, et al. Risks of breast or ovarian cancer in BRCA1 or BRCA2  
703 predictive test negatives: findings from the EMBRACE study. *Genetics in Medicine* (2018)  
704 **20**:1575–1582. doi:10.1038/gim.2018.44
- 705 87. Hu N, Wang C, Han XY, He LJ, Tang ZZ, Giffen C, Emmert-Buck MR, Goldstein AM,  
706 Taylor PR. Evaluation of BRCA2 in the genetic susceptibility of familial esophageal cancer.  
707 *Oncogene* (2004) **23**:852–858. doi:10.1038/sj.onc.1207150
- 708 88. Akbari MR, Malekzadeh R, Nasrollahzadeh D, Amanian D, Islami F, Li S, Zandvakili I,  
709 Shakeri R, Sotoudeh M, Aghcheli K, et al. Germline BRCA2 mutations and the risk of  
710 esophageal squamous cell carcinoma. *Oncogene* (2008) **27**:1290–1296.  
711 doi:10.1038/sj.onc.1210739
- 712 89. Skoulidis F, Cassidy LD, Pisupati V, Jonasson JG, Bjarnason H, Eyfjord JE, Karreth FA,  
713 Lim M, Barber LM, Clatworthy SA, et al. Germline Brca2 Heterozygosity Promotes  
714 KrasG12D -Driven carcinogenesis in a murine model of familial pancreatic cancer. *Cancer  
715 Cell* (2010) **18**:499–509. doi:10.1016/j.ccr.2010.10.015
- 716 90. Arora A, Agarwal D, Abdel-Fatah TMA, Lu H, Croteau DL, Moseley P, Aleskandarany MA,  
717 Green AR, Ball G, Rakha EA, et al. RECQL4 helicase has oncogenic potential in sporadic  
718 breast cancers. *Journal of Pathology* (2016) **238**:495–501. doi:10.1002/path.4681
- 719 91. Zhao L. Mismatch repair protein expression in patients with stage II and III sporadic  
720 colorectal cancer. *Oncology Letters* (2018) **15**:8053–8061. doi:10.3892/ol.2018.8337
- 721 92. Vogelsang M, Wang Y, Veber N, Mwapagha LM, Parker MI. The cumulative effects of  
722 polymorphisms in the DNA mismatch repair genes and tobacco smoking in oesophageal  
723 cancer risk. *PLoS ONE* (2012) **7**:1–10. doi:10.1371/journal.pone.0036962
- 724 93. Haraldsdottir S, Hampel H, Tomsic J, Frankel WL, Pearlman R, de La Chapelle A, Pritchard  
725 CC. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic,  
726 rather than germline, mutations. *Gastroenterology* (2014) **147**:1308-1316.e1.  
727 doi:10.1053/j.gastro.2014.08.041

- 728 94. Fahed AC, Wang M, Homburger JR, Patel AP, Bick AG, Neben CL, Lai C, Brockman D,  
 729 Philippakis A, Ellinor PT, et al. Polygenic background modifies penetrance of monogenic  
 730 variants for tier 1 genomic conditions. *Nature Communications* (2020) **11**:1–9.  
 731 doi:10.1038/s41467-020-17374-3
- 732 95. Werness BA, Parvatiyar P, Ramus SJ, Whittemore AS, Garlinghouse-Jones K, Oakley-  
 733 Girvan I, DiCioccio RA, Wiest J, Tsukada Y, Ponder BAJ, et al. Ovarian carcinoma in situ  
 734 with germline BRCA1 mutation and loss of heterozygosity at BRCA1 and TP53. *Journal of*  
 735 *the National Cancer Institute* (2000) **92**:1088–1091. doi:10.1093/jnci/92.13.1088

736

737

## 738 11 Figure Legends

739 **Figure 1.** The frequency and distribution of cancer susceptibility genes (CSGs) with more than one  
 740 pathogenic/likely pathogenic germline mutation detected in patients with esophageal squamous cell  
 741 cancer (ESCC). Only tumor suppressor genes with more than five mutations are shown. Upper bars  
 742 represent the cumulative mutation numbers of each sample. Bottom bars represent the clinical  
 743 information (race, gender, age, and survival/death) about the patients. The left table presents the  
 744 frequency of mutations shown in the noncancer Genome Aggregation Database (gnomAD) and the  
 745 China Metabolic Analytics Project (ChinaMAP) database. Right bars represent the mutation counts.  
 746 The classification of the CSG is next to the mutation name (gene name + reference SNP number or  
 747 gene name + chromosome position + nucleotide change).

748 **Figure 2.** The distribution of pathogenic/likely pathogenic germline mutations, somatic mutations,  
 749 and allele loss somatic copy number variations (SCNVs) in esophageal squamous cell cancer (ESCC)  
 750 cases with potential double-hit events. Upper bars represent the clinical information (age and race)  
 751 about those patients. Squares represent somatic mutations, triangles represent germline mutations,  
 752 and circles represent allele loss SCNVs.

753 **Figure 3.** The two types of double-hit events. **(A)** The paradigm of double-hit events. **(B)** The  
 754 correlation between age and double-hit event type in esophageal squamous cell cancer (ESCC) cases.  
 755 The position of line is the median age, and the position of rhombus is the mean age in specific ESCC  
 756 cohorts. The digits in the boxes are the numbers of ESCC cases in each category.

757 **Figure 4.** Significantly enriched pathways and networks in esophageal squamous cell cancer  
 758 (ESCC). **(A)** The network composed of genes involved in the top 10 pathways in the Kyoto  
 759 Encyclopedia of Genes and Genomes pathway enrichment. The red dots represent genes, and the blue  
 760 circles represent pathways. The larger the area, the higher the degree of enrichment. The different  
 761 lines represent various categories of pathways; green lines indicate genetic information processing,  
 762 and purple lines indicate human disease. Solid-line rectangles and solid rectangles in blue, red, and  
 763 yellow represent various gene classifications. **(B)** The x axis represents cancer susceptibility genes  
 764 mutated in the Fanconi anemia pathway; the y axis represents the number of patients affected in our  
 765 cohort. Red font: tumor-suppressor genes.

766 **Table 1.** Significance of *TP53*, *BRCA2*, and *RECQL4* pathogenic or likely pathogenic variants for  
 767 ESCC risk in Chinese patients

Gene	Chinese ESCC cohort(n=424)		1000 Genomes EAS(n=504)		$P^3$	OR	95%CI	ChinaMAP (n=10,558) <sup>2</sup>		$P$	OR	95%CI
	$P_{\text{banden}}$	Cases <sup>1</sup> (n = 424)	Controls (n = 504)	Controls (n = 10,558)								

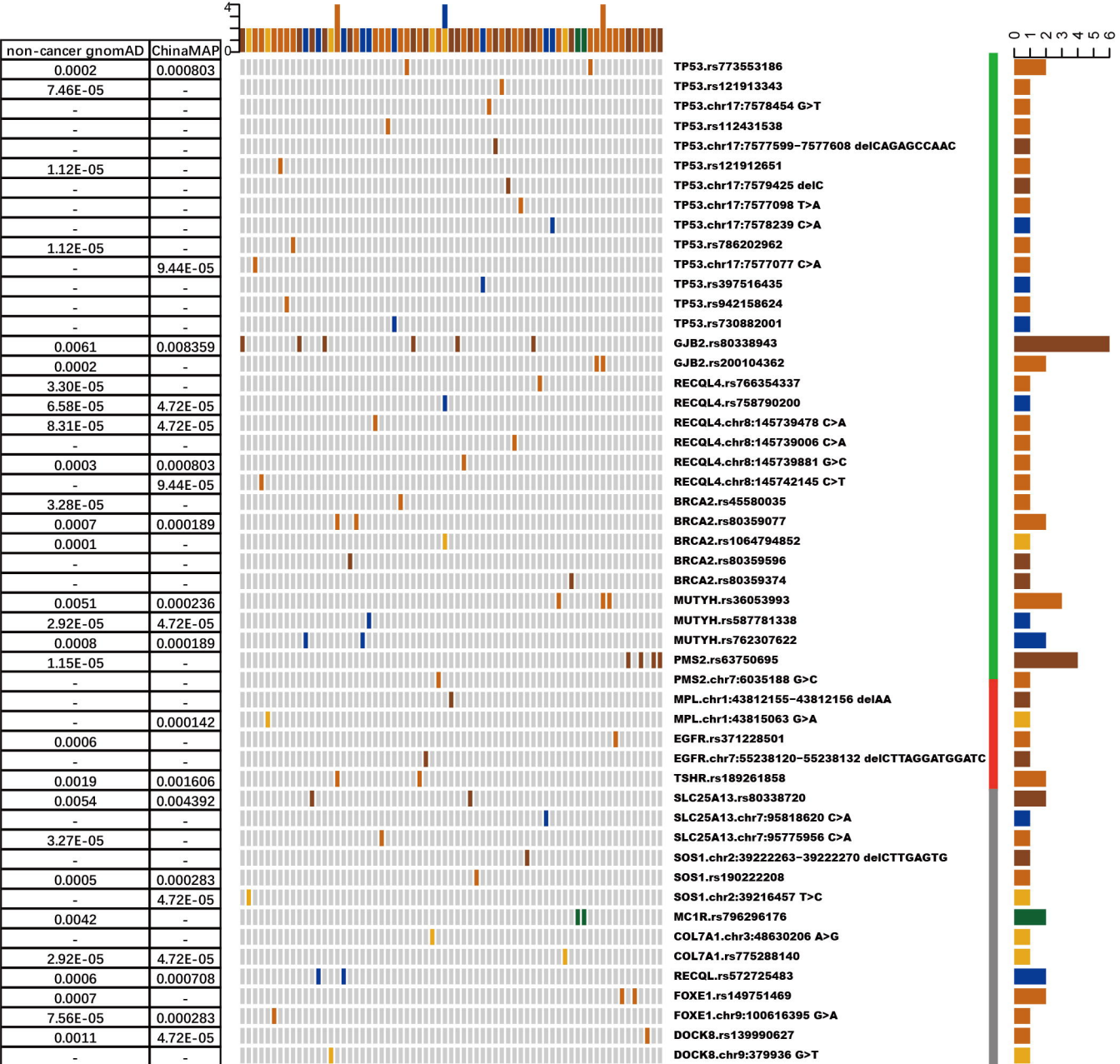
<i>TP53</i>	$3.050 \times 10^{-3}$	14 (3.30%)	4 (0.79%)	$7.359 \times 10^{-3}$	4.26	1.33 to 17.91	34 (0.32%)	$1.851 \times 10^{-9}$	10.59	5.21 to 20.45
<i>BRCA2</i>	0.015	5 (1.18%)	0 (0%)	0.0197	Inf	1.09 to Inf	47 (0.44%)	0.0489	2.68	0.83 to 6.75
<i>RECQL4</i>	0.035	6 (1.14%)	1 (0.20%)	0.0519	7.21	0.87 to 332.23	41 (0.39%)	0.0089	3.69	1.27 to 8.81

768 Abbreviations: ChinaMAP, China Metabolic Analytics Project; EAS, East Asian; ESCC, esophageal  
769 squamous cell cancer; OR, odd ratio; CI, confidence interval; Inf, infinity.

770 <sup>1</sup>Mutation annotation are based on *TP53* transcript: NM\_001126112, *BRCA2* transcript: NM\_000059  
771 and *RECQL4* transcript: NM\_004260.

772 <sup>2</sup>ChinaMAP, *TP53*, *BRCA2*, and *RECQL4* variants were exported from <http://www.mbiobank.com/>  
773 on June 2, 2020.

774 <sup>3</sup>Fisher exact test.



**Type**

- Frameshift deletion
- Frameshift insertion
- Missense
- Splice\_site
- Nonsense

**Race**

- Asian
- Caucasian
- Black or African American

**Gender**

- Male
- Female

**Age**

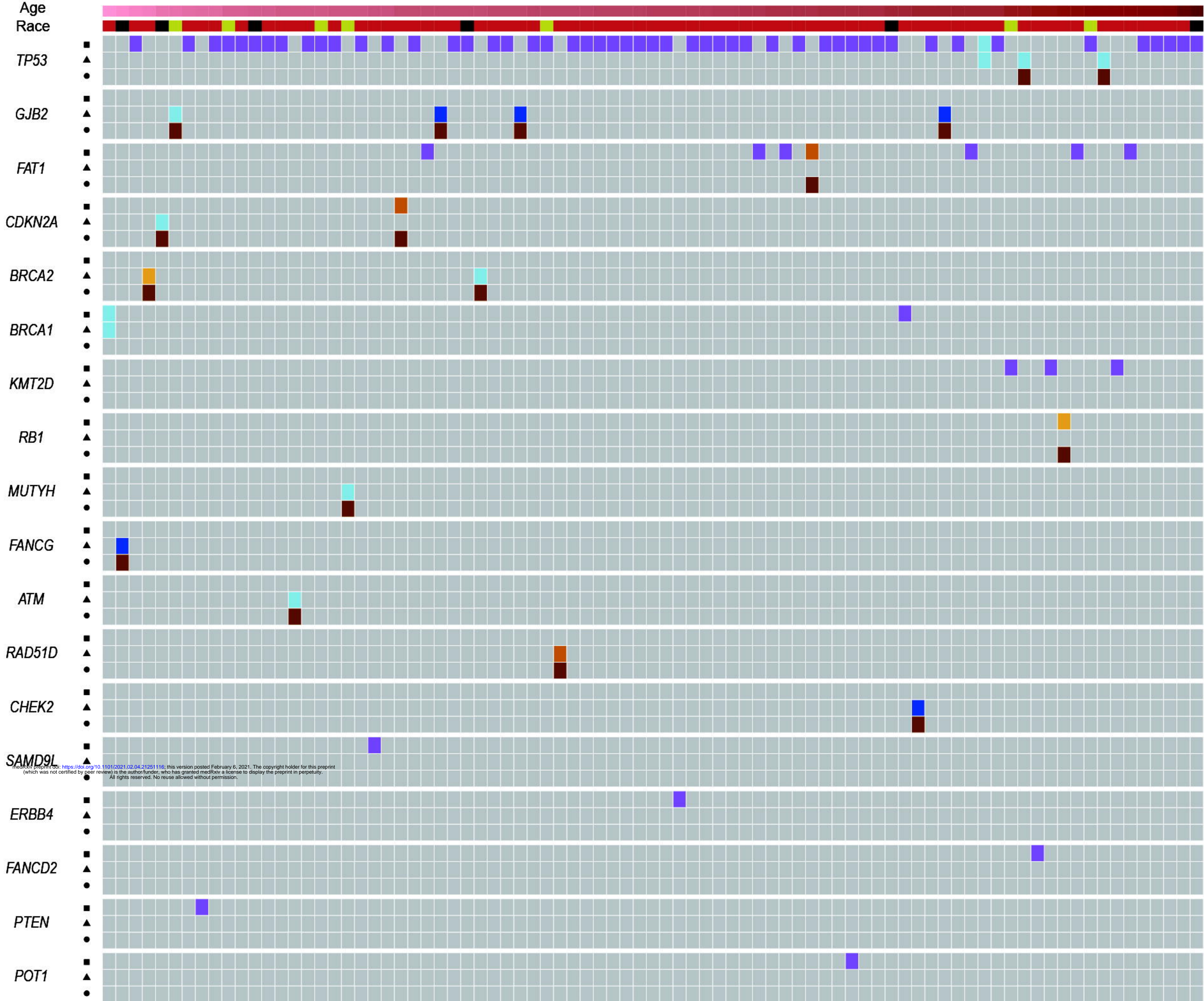
100  
80  
60  
40  
20

**Death/Survival**

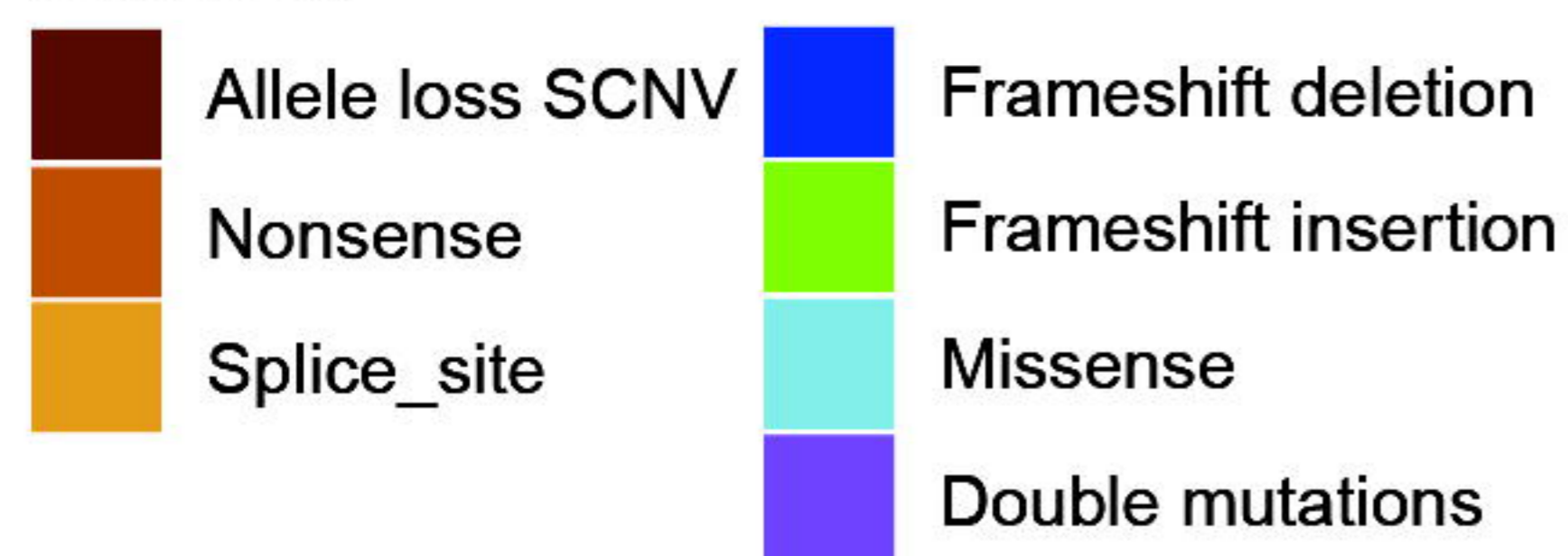
- Alive
- Dead
- NA

**Gene classification**

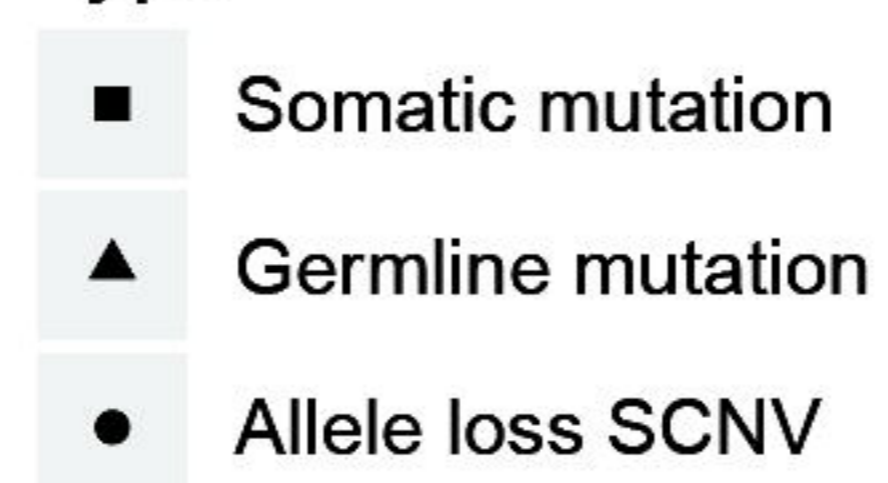
- Tumor suppressor gene
- Oncogene
- Nonclassified gene



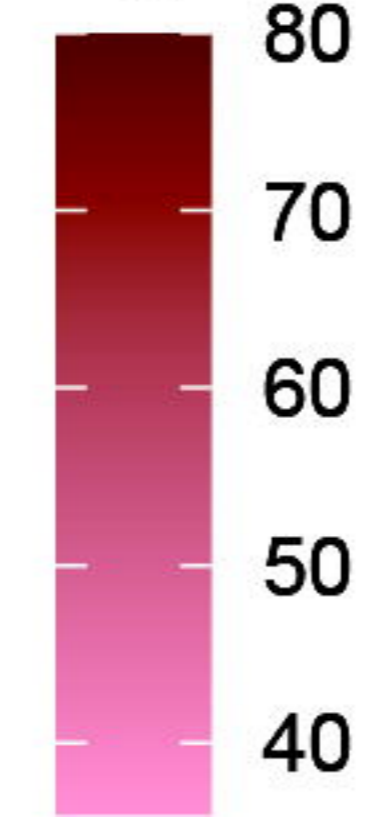
### Mutation



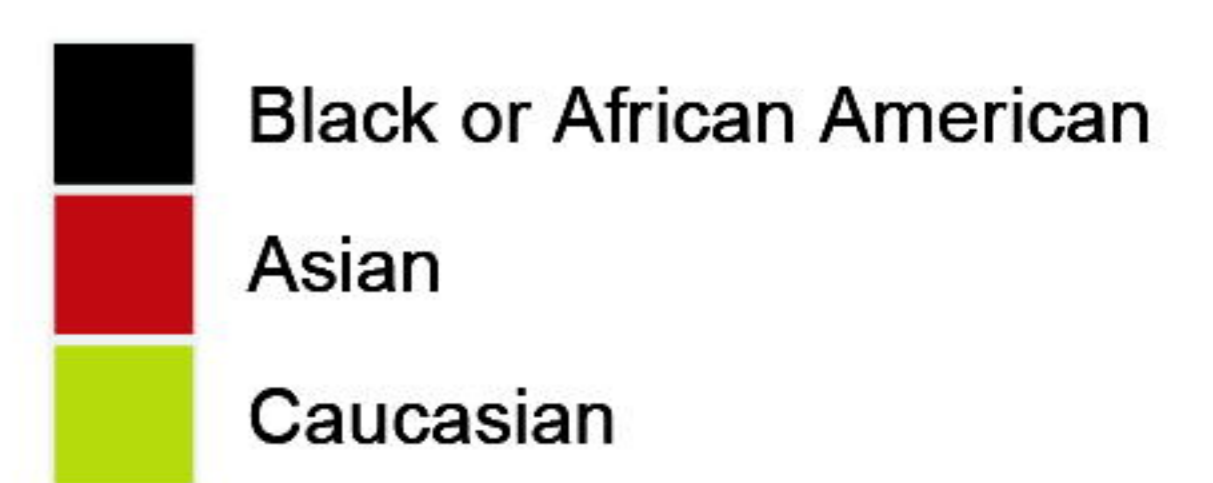
### Type



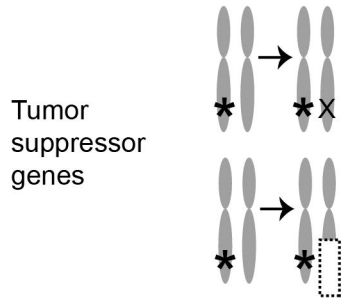
### Age



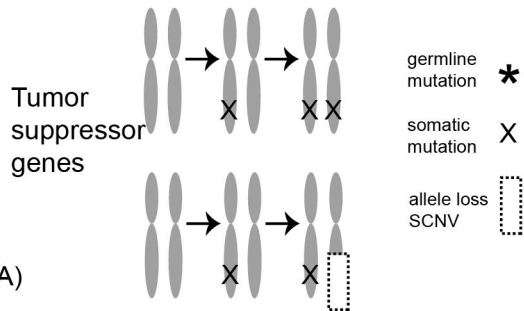
### Race



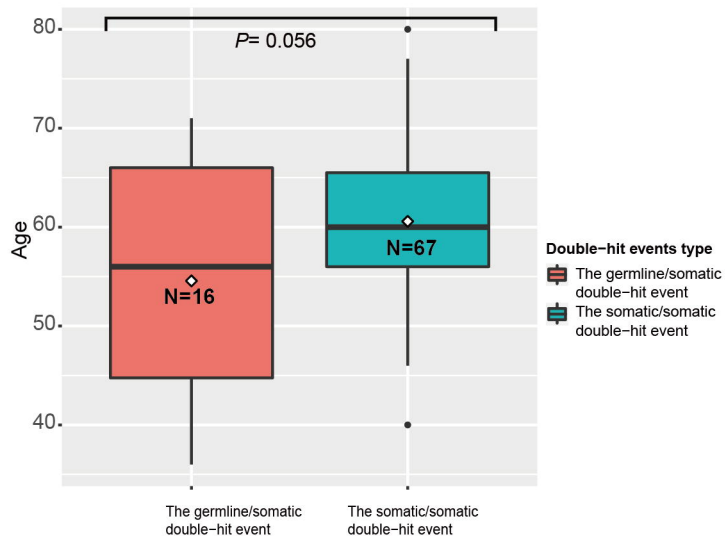
## The germline/somatic double-hit events



## The somatic/somatic double-hit events



(A)



(B)

