

# A Systematic Literature Review of Whole Exome and Genome Sequencing Population Studies of Genetic Susceptibility to Cancer

Melissa Rotunno<sup>1</sup>, Rolando Barajas<sup>1</sup>, Mindy Clyne<sup>1</sup>, Elise Hoover<sup>1</sup>, Naoko I. Simonds<sup>2</sup>, Tram Kim Lam<sup>1</sup>, Leah E. Mechanic<sup>1</sup>, Alisa M. Goldstein<sup>1</sup>, and Elizabeth M. Gillanders<sup>1</sup>

## ABSTRACT

The application of next-generation sequencing (NGS) technologies in cancer research has accelerated the discovery of somatic mutations; however, progress in the identification of germline variation associated with cancer risk is less clear. We conducted a systematic literature review of cancer genetic susceptibility studies that used NGS technologies at an exome/genome-wide scale to obtain a fuller understanding of the research landscape to date and to inform future studies. The variability across studies on methodologies and reporting was considerable. Most studies sequenced few high-risk (mainly European) families, used a candidate analysis approach, and

identified potential cancer-related germline variants or genes in a small fraction of the sequenced cancer cases. This review highlights the importance of establishing consensus on standards for the application and reporting of variants filtering strategies. It also describes the progress in the identification of cancer-related germline variation to date. These findings point to the untapped potential in conducting studies with appropriately sized and racially diverse families and populations, combining results across studies and expanding beyond a candidate analysis approach to advance the discovery of genetic variation that accounts for the unexplained cancer heritability.

## Introduction

Since 2005, the volume of publications enabled by sequencing approaches has grown at an astonishing rate. Several reviews have described the sequencing technology platforms (1) and advancements made in next-generation sequencing (NGS) over the past decade (2). The widespread availability of NGS technologies, including whole genome sequencing (WGS) and whole exome sequencing (WES), has not only led to its applications in cancer research, but also for use in the clinical setting (3, 4).

Use of NGS has accelerated the discovery of somatic mutations (5) and germline mutations in Mendelian diseases (6). Approximately 60% of Mendelian disease projects have successfully identified disease gene mutations (7) using sequencing technologies, improving upon classical approaches for gene discovery (e.g., linkage analysis). In addition, the application of NGS technologies is revealing complex somatic mutational signatures associated with different types of cancer, a disease that is, by definition, a result of somatic mutations (8, 9).

Given these successes, there is hope that sequencing studies may also aid the identification of genes accounting for the expected cancer heritability. The estimated cancer heritability is 33% for overall cancer (10) and its unexplained component has remained high, where a study by Susswein and colleagues (11) reported that 91% of cancer cases tested negative for known mutations in a large gene-panel testing study. In addition to the heritability hidden in current array-based studies and likely detectable with larger sample sizes, it has been hypothesized that the missing familial heritability may reside in rare variants of high or moderate/low penetrance that are potentially tractable by NGS technologies (12). As NGS technologies continue to evolve, NGS will play an increasing role in cancer research in the foreseeable future.

We conducted a systematic literature review and evaluated the degree of success and limitations in identifying germline cancer susceptibility variants using NGS technologies at a genome-wide scale, that is, through WES and WGS, with the goal of learning from past efforts and obtaining a fuller understanding of the NGS-related research landscape to date. Given the transition of genomic discovery research from candidate genes (historically of limited success) to WES/WGS studies, the high cost of WES/WGS methods, and their specific challenge with sifting through millions of variants, this review focuses on the effectiveness of WES/WGS studies, and not on other NGS gene-targeted approaches, to identify novel variants and genes involved in cancer risk. This review provides selected study-related characteristics, technologies, and methodologic details for 186 WES/WGS-related publications with the goal of informing the design of future studies. We also discuss the research needs and opportunities that could further advance the discovery of cancer susceptibility genes or variants. It should be noted that, although the reviewed articles were not selected on the basis of their focus on rare versus common variants nor on their focus on low versus moderate/high penetrant variants, more cost-efficient approaches, based on genome-wide genotyping assays, exist to study common variants, while NGS technologies are necessary to study rare variants.

<sup>1</sup>National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services, Bethesda, Maryland. <sup>2</sup>Scientific Consulting Group, Inc., Gaithersburg, Maryland.

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

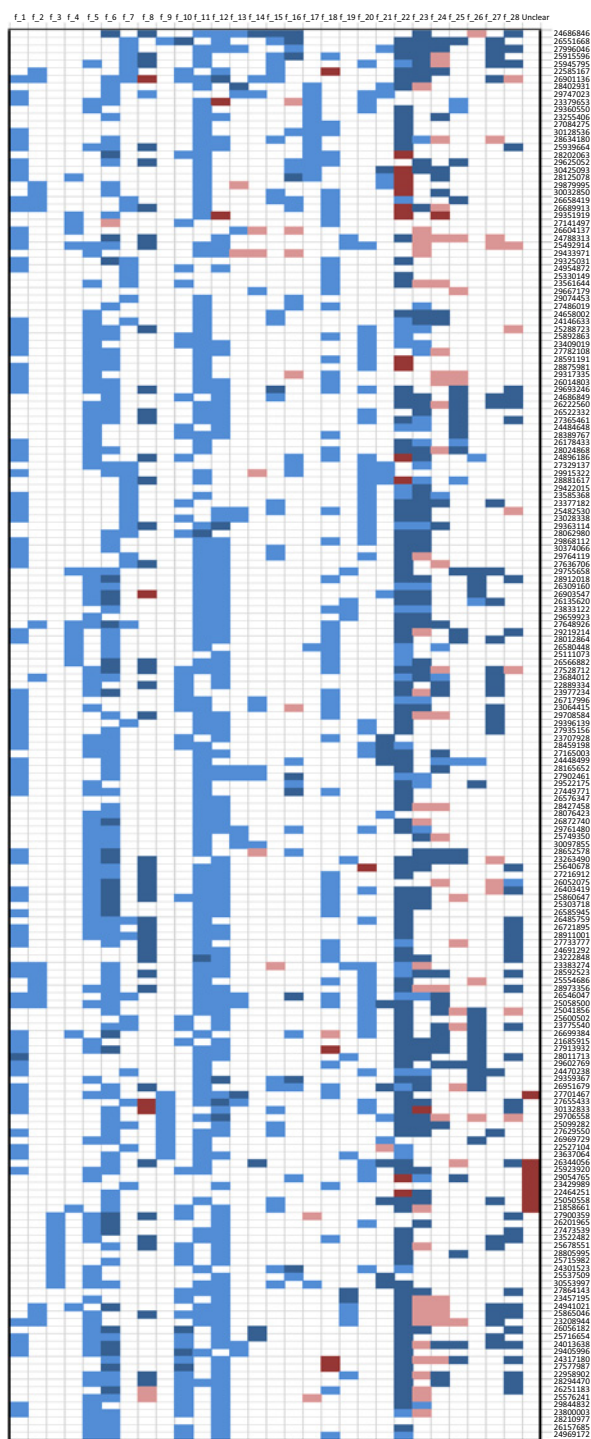
A.M. Goldstein and E.M. Gillanders contributed equally as co-senior authors of this article.

**Corresponding Author:** Melissa Rotunno, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892. Phone: 240-276-7245; Fax: 301-402-4489; E-mail: [rotunnom@mail.nih.gov](mailto:rotunnom@mail.nih.gov)

Cancer Epidemiol Biomarkers Prev 2020;29:1519-34

doi: 10.1158/1055-9965.EPI-19-1551

©2020 American Association for Cancer Research.



**Figure 1.** Summary across the 186 reviewed articles of the methods used to conclude if a variant or gene was plausibly involved in a causal pathway to cancer. The y-axis displays the PMID of the 186 coded articles; the x-axis displays 29 broad categories of filtering criteria described in **Table 1**. White color indicates that the criterion was not used by the authors to identify a variant/gene as possibly linked to the cancer under investigation. Light blue indicates that the criterion was used as a selection filter. Dark blue indicates that the criterion was used as increased evidence of variant/gene-cancer association. Pink color indicates that

## Methodology

We followed the methodology for systematic literature review according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (13).

### Search strategy

For this report, PubMed and Embase were searched for the period between January 1, 2000 through December 31, 2018, using various search terms for exome/genome sequencing and germline susceptibility and cancer (see Supplementary Fig. S1). Both searches were restricted to articles written in English. The references of the reviewed articles were also checked for the presence of additional in scope articles that may have been missed by the above keyword searches.

### Exclusion/inclusion criteria

Articles were included in this literature review if they had used genome wide (i.e., whole exome or whole genome) sequencing to generate germline DNA data in at least two cancer cases (even if within the same family) with the purpose of identifying cancer susceptibility genes or variants (see exclusion criteria in Supplementary Fig. S1). We excluded articles that sequenced only one cancer case because they were often case reports without a variant/gene identification research focus. Notably, we did not exclude publications that restricted the analysis to *a priori* or candidate genetic regions (referred in this manuscript as “candidate analysis approach”), due to the large methodologic diversity across such studies; instead, we opted to include all studies that used WES/WGS independent of their use of a candidate analytic approach, and captured the prior knowledge used to select variants/genes to allow for sensitivity analysis by candidate analysis approach. The eligibility of each abstract/full-text article was assessed independently in a standardized manner by three reviewers. A fourth reviewer confirmed all inclusions and performed quality control on one third of the exclusions. The exclusion criteria were applied in no particular order, and the first reason noted by the coder was recorded, even if an article could be excluded for multiple reasons.

### Data abstraction

We looked for four broad components or phases in each article: (i) a required discovery phase and optional phases of (ii) technical validation, (iii) independent replication, or (iv) functional evaluation.

- (i) The discovery phase refers to the component of the article where germline DNA of more than one cancer case (and possibly controls) were sequenced by WES or WGS with the goal of identifying cancer susceptibility variants and genes.
- (ii) The technical validation attempts to confirm some of the variants observed in the discovery phase using an alternative sequencing technology.

the criterion was used as decreased evidence of variant/gene-cancer association. Dark red indicates that the information related to that criterion was unclear. The last category on the right, “unclear,” indicates that some other not clearly stated criteria were used. The order of the articles along the y-axis is based on a computed correlation between values 2 (for supporting evidence), 1 (for used selection criteria), 0 (for not used criteria), -1 (for unsupporting evidence), and -2 (for unclear), that is, articles using similar filtering criteria sets are shown next to each other.

- (iii) The independent replication phase attempts to replicate some of the variants or genes observed in the discovery phase in independent cases.
- (iv) The functional evaluation phase characterizes through *in silico* and/or laboratory functional experiments some of the variants or genes observed in the discovery phase.

National Comprehensive Cancer Network (NCCN) guidelines criteria for genetic cancer risk assessment. Of note, we defined “familial studies” as those studies that sequenced samples from cancer cases (here referred to as “familial cases”) belonging to a family in which multiple cancer cases of the studied cancer type had been diagnosed (also referred to in the literature as “high-risk families”). We defined “unselected studies” as those studies that sequenced samples from cancer cases (here referred to as “unselected cases”) who were unselected for family history of cancer (also referred to in the literature as “sporadic cases”). For each authors' selected variants and genes, we recorded the nomenclature and minor allele frequency (MAF), as reported by the authors, the number of families for which cancer status fully segregated or not with the variant, the number of unselected or familial cases, unaffected relatives, or unrelated controls that carried

Data abstracted (see list and definitions of coding fields in **Table 1**; Supplementary Table S1) included publication and general study information, numbers, and characteristics of cases, controls, and families used in each phase, sequencing technique, data filtering and analysis methods, *in silico* and experimental functional assessment, and key conclusions. We extracted information on family history of cancer, early age at diagnosis and/or multiple primaries, which are the

**Table 1.** Coding fields used for reviewing articles.

Topic	Coding fields
Publication	Pubmed ID; Journal; Year; Author; Title; Abstract
Study general	Goal; Study design; Source of individuals; Cancer type; Ethnicity; Sequencing center; Data repository;
Number of individuals sequenced in discovery	Number of cases, controls, families, and cases per family sequenced in discovery phase
Sequencing technique	Samples type; Exome and/or genome; Capture kit; Sequencer; Coverage/depth
Processing of raw data	Aligner; Reference genome; Variant caller and calling quality control; Annotation software and sources
Technical validation	Yes/no; Validation technology; Number of cases, controls, families, and cases per family sequenced in validation phase; variants/genes validated
Independent replication	Yes/no; Replication technology and analysis; Number of cases, controls, families, and cases per family sequenced in replication phase; variants/genes replicated
Functional validation	<i>In silico</i> functional analyses; Experimental functional study
Variants and genes data analysis	Candidate analysis approach; Filtering strategy overall; Analytical methods
Variants and genes identification	Yes/no; Identified variants and genes; Number of cases, controls, families carrying the identified variants and genes
Authors comments and conclusions	Challenges; Suggested next steps; Conclusions
Derived filtering criteria/categories shown in <b>Fig. 1</b>	f_1: Variant passing quality control metrics f_2: Heterozygous variant f_3: Homozygous variant f_4: Variant located in a coding region f_5: Nonsynonymous or splice variant f_6: Variant damaging according to <i>in silico</i> algorithms f_7: Truncating variant f_8: Variant altering protein properties according to molecular modeling f_9: Not hypervariable gene f_10: Variant absent from minor allele frequency (MAF) databases f_11: Variant rare in MAF databases f_12: Variant segregating with disease status in the family f_13: Variant present in multiple families or independent cases f_14: Variant enriched in cases compared to controls f_15: Gene mutated in multiple families or independent cases f_16: Gene enriched in cases compared to controls f_17: Variant present in disease-related databases f_18: Gene known to be linked to disease f_19: Genetic region known to be linked to disease f_20: Biological or molecular pathway known to be linked to disease f_21: Pathway analysis indicating a gene-disease link f_22: Variant confirmed through technical validation f_23: Variant or gene replicating in independent cases f_24: Variant loss of heterozygosity (LOH) observed in tumor f_25: Relevant somatic mutations observed in tumor f_26: Gene-disease link supported by functional experiment f_27: Variant splicing supported by experiment f_28: Variant-disease link supported by functional experiment

Downloaded from <http://aacrjournals.org/cebp/article-pdf/29/8/1519/1945786/1519.pdf> by guest on 22 June 2024

the variant, and the number of families, familial cases, unselected cases, or controls carrying any prioritized variant in the same gene. For quality control purposes, 50% of the articles were also reviewed by a second abstractor and discrepancies between coders were resolved by consensus. Information in the articles that was unclear was coded on the basis of best interpretation of the reviewers.

## Results

### Article selection

The search yielded a total of 6,339 unique articles (see PRISMA flowchart in Supplementary Fig. S1) that were evaluated for inclusion and exclusion criteria (Supplementary Table S2). After full-text review, 186 articles met the inclusion criteria and are listed with the derived coding variables in Supplementary Tables S1 and S3. The distribution of the 186 reviewed articles by publication year shows an increase from 2 to 40 articles per year between 2011 and 2015, followed by a plateau in 2016 to 2018 (Supplementary Fig. S2).

### Study design and population characteristics

In the discovery phase, 86% of articles used familial cases (11% of which were in combination with early age of onset and/or unselected cases), 12% of studies were conducted in unselected cases, and 2% of articles used early age of onset cases. Fifty-five percent of the studies included controls in the discovery phase, of which 67% were unaffected relatives of the cancer cases. Fifty-seven percent of the reviewed articles also attempted some type of replication in an independent group of cancer cases. However, only 17% of the replication phase used the same study design as the discovery phase (Supplementary Fig. S3). For example, 16% of the family-based articles that included a replication phase did not include familial cancer cases. Moreover, controls were included more often in the replication phase (80%) than in the discovery phase (55%). Likewise, unselected cases were used more frequently in the replication phase than in the discovery phase (30% compared with 12%).

One third of the studies sequenced only 2 or 3 cancer cases in the discovery phase and only 28 (15%) of the articles sequenced more than 100 cancer cases. Technical validation of some of the variants was performed in 85% of the articles, although most of these were conducted in studies with small sample size (i.e., less than 50 cases). The replication phase had generally larger sample size, including more than 100 cancer cases. Almost half ( $n = 73$ ) of the 160 family-based studies sequenced a single family and the majority ( $n = 115$ ) sequenced one or two familial cancer cases (Table 2).

The most commonly studied cancer types were breast cancer (15%), followed by hematologic malignancies (15%, which included pediatric cases), colorectal cancer (10%), melanoma (7%), lung cancer (7%), and prostate cancer (5%; Supplementary Fig. S4). Information on race, ethnicity, or country of origin was reported in 85% of the articles reviewed and mostly referred to the region or country of origin (Supplementary Fig. S5), with only few studies reporting sequencing-derived ancestry. Over half of the studies were conducted in Caucasians or individuals from Europe (59%), followed by individuals from Asia (13%), the Middle East (7%), of African descent (3%), from Latin America (2%) and Australia (2%).

### Sequencing technologies, read alignment, variant calling, and annotation

Sequencing was performed in DNA extracted from blood (65%), formalin-fixed paraffin-embedded, generally nontumor, tissues (8%) and/or saliva (5%). However, a notable proportion did not state the

**Table 2.** Number and percentage of cancer cases and families studied per article and by study phase.

	Number (%) of articles		
	Discovery	Independent replication	Technical validation
<b>Number of cancer cases</b>			
2–3	62 (33%)	0 (0%)	41 (22%)
4–10	38 (20%)	3 (2%)	48 (26%)
11–50	39 (21%)	17 (9%)	35 (19%)
51–100	19 (10%)	10 (9%)	4 (2%)
101–1,000	22 (12%)	43 (23%)	3 (2%)
>1,000	6 (3%)	34 (18%)	1 (1%)
Not stated	0 (0%)	0 (0%)	26 (14%)
Total <sup>a</sup>	186 (100%)	107 (57%)	158 (85%)
<b>Number of high-risk families</b>			
1	73 (39%)	3 (2%)	
2–10	27 (15%)	6 (3%)	
11–50	40 (22%)	23 (12%)	
51–100	10 (5%)	8 (4%)	
101–1,000	6 (3%)	23 (12%)	
>1,000	2 (1%)	5 (3%)	
Not stated	2 (1%)	4 (2%)	
Total <sup>a</sup>	160 (86%)	72 (39%)	
<b>Average number of sequenced cases per family</b>			
1	49 (26%)	56 (30%)	
2	66 (35%)	10 (5%)	
3	27 (15%)	1 (1%)	
4	10 (5%)	1 (1%)	
5	4 (2%)	0 (0%)	
6–7	3 (2%)	0 (0%)	
Not stated	1 (1%)	5 (2%)	
Total <sup>a</sup>	160 (86%)	72 (39%)	

<sup>a</sup>Some totals do not add to 100% because 15% of articles did not perform technical validation; 43% of articles did not perform independent replication; 14% of articles did not include familial cases in the discovery phase; and 61% of articles did not perform independent replication in familial cases.

DNA source (22%). Moreover, the amount of DNA used for sequencing (~1–3 µg per sample) was reported only in 33% of the articles (Supplementary Table S1).

Ten articles (5%) analyzed WGS only, seven studies (4%) conducted both WGS and WES, and 169 studies (91%) analyzed WES only. The reviewed studies used 28 different capture methods and 14 different sequencing platforms (Supplementary Fig. S6), whereas 12% and 8% did not report the capture or sequencer used, respectively. Sequencing coverage information was reported for only 71% of the articles, as average depth (52%) and/or percentage of the target genome covered at 10× or higher thresholds (42%; Supplementary Fig. S7). For most articles, it was unclear whether the reported coverage statistics referred to the targeted or actual coverage and to pre-quality control or post-quality control coverage. In addition, there was no correlation between the number of samples sequenced and the reported coverage depth.

Sequencing reads were aligned to human genome references, where 79% of the studies reported using hg19 (also known as NCBI build 37 or GRCh37) and 3% used hg18 (a.k.a., NCBI build 36). The most widely used aligner (52%) was the Burrows-Wheeler Aligner (BWA; ref. 14). Reference genome and alignment algorithm used were not specified in 18% and 12% of articles, respectively. Over half of the articles used Genome Analysis Toolkit (GATK; ref. 15) variant calling algorithms. One fourth of the studies used more than one algorithm to

call variants, which can improve call quality, and 8% of articles did not report their variant calling method. Various quality metrics were applied to screen the sequencing reads (e.g., removal of PCR duplicates, unmapped or nonuniquely mapped, or out of target reads) and the called variants (e.g., removal of variants with quality or coverage below a preset threshold; Supplementary Table S1). In addition, 26% of the articles used additional control sequencing datasets generated in-house via the same technology and pipeline as the study dataset to control for technical artifacts.

The annotation software used to annotate the called variants was not specified in 30% of the articles and included ANNOVAR (16) in 36% of the articles. Eighty-nine percent of the articles used allele frequency information from publicly available databases, mainly from the 1000 Genomes Project (ref. 17; 64%), the NHLBI Exome Sequencing Project (ESP; ref. 18; 47%), dbSNP (ref. 19; 46%), and the Exome Aggregation Consortium (ExAC; ref. 20; 32%). Several other annotation tools and source databases were used (Supplementary Table S1).

### Criteria used to filter variants and genes

Figure 1 shows which criteria or filters (see the  $f_1, \dots, f_{28}$  variables described in Table 1) were used in each reviewed article. Filtering criteria were generally used to prioritize/select a variant/gene over others and/or to seek evidence in support of a selected variant/gene. Data in Fig. 1 show that the criteria used to identify variants and genes with a role in cancer susceptibility are disparate across articles, and consequently, results cannot be directly compared across studies. Below, we examine these filtering/selection criteria and summarize their use and outcomes by grouping them into seven broader themes: (i) variant quality; (ii) variant effect; (iii) variant rarity; (iv) mode of inheritance and genetic disease association; (v) candidate analysis approaches; (vi) independent replication; and (vii) functional validation. Two general observations hold true: (i) lack of sensitivity analyses to assess variability in results by changes in variant/gene selection strategy, and (ii) no or minimal reporting of a justification for the choice of criteria and thresholds used.

- (i) Variant quality ( $f_1, f_9, f_{22}$ ). Approximately half of the reviewed articles explicitly described the use of variant call metrics to exclude low-quality variants from the analyses, for example, manual inspection using the Integrative Genomics Viewer (21), removing variants in paralogs or repeats regions, and/or with Phred-scaled quality scores or coverage below a given threshold (Supplementary Table S1). Approximately 85% of the articles reported technically validating variants, where most used Sanger sequencing. The technical validation success rate was about 80% for studies that tested over 50 variants versus above 90% for studies that tested fewer variants (Supplementary Fig. S8).
- (ii) Variant effect ( $f_4, f_5, f_6, f_7, f_8$ ). Most articles ( $n = 168$  or 90%) required the variants to be in coding regions, and more specifically, to be nonsynonymous or in splice sites or frameshift ( $n = 163$  or 88%). A subset of these articles also required the selected variants to be functionally impactful (e.g., “deleterious”, “damaging”, or “pathogenic”) according to various *in silico* algorithms ( $n = 91$  or 49%) or to be truncating ( $n = 38$  or 20%). Supplementary Table S1 lists for each article the adopted *in silico* pathogenicity predictors (e.g., refs. 22–29) that were often used in combination.
- (iii) Variant rarity ( $f_{10}, f_{11}$ ). Most articles required the selected variants to be absent/not described (25%) or rare (62%) based on a preset MAF threshold (0–0.1 range, Supplementary Fig. S9) in

internal or publicly available control datasets, such as 1000 Genomes Project, dbSNP, ESP, or others (Supplementary Table S1).

- (iv) Mode of inheritance ( $f_2, f_3, f_{12}$ ) and genetic disease association ( $f_{13}, f_{14}, f_{15}, f_{16}, f_{21}$ ). Only 10 (5%) and 17 (9%) articles restricted their search to homozygous and heterozygous variants according to a recessive or a dominant mode of inheritance, respectively. Because the remaining articles describe only heterozygous variants in their findings, a dominant inheritance hypothesis can be assumed for all but 10 articles. The majority of articles ( $n = 108$  or 58%) required the variant to fully or partially segregate with disease status in at least one family whose members were sequenced in the discovery phase. Only two of the reviewed articles looked for *de novo* variants. In several studies, the selected genes were required to be mutated in more than one family ( $n = 19$  or 10%) or in multiple independent cases ( $n = 5$  or 3%), or to be enriched in cases compared with controls according to burden tests ( $n = 18$  or 10%). Fewer articles required the same selected variant to be present in more than one family ( $n = 10$  or 5%), in multiple independent cases ( $n = 6$  or 3%), or to be statistically enriched in cases compared with controls ( $n = 8$  or 4%). Finally, 12 articles used pathway analysis techniques to identify biological or molecular functions that were enriched with mutated genes.
- (v) Candidate analysis approaches ( $f_{17}, f_{18}, f_{19}, f_{20}$ ). Two thirds of the reviewed articles ( $n = 121$ , 65%) used existing information from the literature and curated databases to restrict the discovery analysis to: variants present in disease-related databases such as ClinVar (ref. 30;  $n = 14$  or 7%); genes known to be linked to disease such as those listed in OMIM (31) or reported in the literature ( $n = 59$  or 32%); genetic region known to be linked to disease through genome wide association studies (GWAS) or linkage studies ( $n = 14$  or 7%); and/or biological or molecular pathways known to be linked to disease, such as DNA repair pathways ( $n = 51$  or 27%).
- (vi) Independent replication ( $f_{23}$ ). Only 107 (57%) articles attempted replication of variants/genes in an independent set of cancer cases. Overall, 79 (42%) reported various degrees of confirmatory evidence. In some cases, the authors reported the presence of other pathogenic variants in the same gene, whereas in others, a statistically significant burden test in cases compared with controls for that gene was reported. In a few studies, the exact variant(s) initially found in the discovery phase were found in additional cancer cases in the replication phase.
- (vii) Functional validation ( $f_{24}, f_{25}, f_{26}, f_{27}, f_{28}$ ). In those studies that evaluated function of the identified variants/genes (70% of 186), 60 (32%) tested for loss of heterozygosity in tumor samples (58% of which tested positive); 36 (19%) looked for somatic mutations in the same gene (69% of which were found); 22 (12%) looked for gene/methylation expression changes supporting a link with disease (86% with positive results); 35 (19%) checked for variant splicing (86% of which were verified); and 45 (24%) carried out *in vitro* experiments or other functional assays on the identified variants, 80% of which showed results consistent with the hypothesized function for these variants.

### Variants and genes identified

About 95% ( $n = 176$ ) of reviewed articles indicated that they identified variants or genes (listed in Table 3 with PMIDs by cancer type) with various degrees of certainty. Only eight (4%) articles clearly

**Table 3.** List and counts of identified genes and PubMed IDs by cancer type.

Cancer type	Authors' identified genes <sup>a</sup> [or pathways]	# genes [unique]	PMID <sup>b</sup> (references) {number of articles}
Biliary	<b>BRCA1, BRCA2, RAD51D, MLH1, MSH2, POLD1, POLE, TP53, ATM,</b>	9 [9]	29360550 (32); {1}
Blood	none; none; <b>ACAN; ATM; CBL; CEBPA; CORO1A; DICER1; ETV6; FAAP24; HAVCR2; ITGB2; KDR; MLL; PAX5; POT1; PRDM9; TP53; TYK2; LAPTM5; HCLS1; SLC26A6, FAM107A; POT1, ACD, TERF2IP; ZXDC, ATNI, LRRC3; FNCP/SLX4, FANCA, GEN1; POLK, PRKCB, ZNF676, PRRC2B, PCDHGB6, GNLSL, TTC36, OTOG, OSGEPL1, RASSF9; MLL3; JAK2, MPL, TYK2, BCL2, BCL3, KIT, SH2B3, MCM3AP, PEAR1, TET2</b>	52 [50]	23379653; 27141497; 25715982; 28652578; 25939664; 26721895; 23522482; 29708584; 26522332; 27473539; 30374066; 27629550; 27365461; 23457195; 24013638; 29693246; 23222848; 23255406; 27733777; 26903547; 29755658; 27528712; 28427458; 26201965; 30097855; 24301523; 30553997; {27}
Bone	<b>ANOS; EXT2; T</b>	3 [3]	27216912; 27636706; 23064415; (60–62) {3}
Brain and nervous system	<b>CASP9; GALNT14; JMJD1C; PMS2; POT1; SMARCE1; SUFU; TP53, ATRX; TP53, MSH4, LATS1; CNL2, ANKRD65, TAS1R2, PINK1, DCAF6, F5, TNN, CFH, DDX59, IGFN1, CHI3L1, CRI, OR2B11, TRIM58, RRM2, ANKRD53, CAPG, IMMT, THNSL2, PIKFYVE, COL4A3, SUMF1, CEPT20, ARAP3, ARHGFEF37, TCOF1, FAT2, OR2V2, HIVEP1, PMS2, ADAM22, C7orf62, TRPA1, TG, ART1, TRIM22, IGSF22, CD6, CD5, SLC22A24, CD248, MYEOV, KIAA1731, PIK3C2G, LRMP, OVCH1, TPPP2, RPGRIP1, TLL5, KCNK10, C14orf159, ACSBG1, ACAN, C15orf42, KIF7, ZNF23, ZNF469, ANKRD11, STARD3, ERBB2, AKAP1, FAM20A, ABCA6, ABCA10, DNAH17, LAMA1, DOPEY2, MYO18B, NEFH, CCDC157, RNF215, SMTN, APOL3</b>	85 [83]	27935156; 26309160; 24896186; 28805995; 25482530; 23377182; 22958902; 29602769; 25041856; 25537509; (63–72) {10}
Breast	none; none; <b>APOBEC3B; ATM; ERCC3; FANCM; GPRC5A; KAT6B; PALB2; RCCT1; RECQL; RECQL; RINT1; RNASEL; XRCC2; BRCA2, STK11; FANCC, BLM; ERCC6, BRCA2, BRCA1; BRCA1, TP53, PTEN, PALB2, RAD51C; XCR1, DLL1, TH, ACCS, SPPL3, CCNF, SRL; BRCA1, BRCA2, PALB2, ATM, CHEK2, BARD1, BRIP1, XRCC2; FANCM, WNT8A, SLBP, CNTRB, AXIN1, TIMP3, PTPRF, UBA3, MAPKAP1, TNFSF8, SIPR3; HSD3B1, CFTR, PBK, ITIH2, MMS19, PABPC3, PPL, DNAH3, LRRC29, CALCOCO2, ZNF677, RASSF2; NOTCH2, DNAH7, RAF1, MST1R, LAMB4, NIN, SLX4, ERCC1, SLC22A16, PTPRD, ARHGFEF2, ERBB2; ABCC12, APC, ATM, BRCA1, BRCA2, CDH1, ERCC6, MSH2, POLH, PRF1, SLX4, STK11, TP53; ABCA10, CHST15, GRIP1, LOC100129697, LOC388813, NBPFF10, PABPC3, C16orf62, KRTAP21–3, NPIP81, PDE4DIP, CCDC7, CFAP46, CXorf23, SMIM13, GAGE2A, PHIP, SLC15A5, ZNF750, ATP10B, PIGN, PRR14L; ABL1, ADRA2A, DUX2, GATA3, GPRIN1, JAKMIP3, KAT6B, LIG1, LIG4, NANP, NFKBIZ, NFRKB, NOTCH2, PHKB, PINK1, POLK, POLQ, PPF1A4, PRKCG, RAD23B, SMG1, TNK2, UBE2L3; ABCC11, ADARB1, ADPRH, ATG4C, ATM, ATXN3L, BANK1, BIRC8, BRX1, BTN2A3, C10orf68, C14orf38, C17orf57, C1orf168, C20orf186, C2orf63, C5orf52, C7orf46, CA12, CASP5, CAT, CCDC99, CDKL2, CFHR5, CGRRF1, CHEK2, CHRN3, CIB1, CPAMD8, CRYBG3, CXCL6, DCD, DCHS2, DDX60L, DNMT3A, ECEL1, EEF2K, EIF2B4, ENDOD1, ENTPD4, EPS8L1, EPST11, ERCC3, EYS, FAM40B, FAP, FETUB, FILIP1, FLG2, FLOT2, FOLH1B, FTMT, GIMAP6, GPRC5A, GTPBP5, HEATR7B2, IFIT2, IGSF22, IL25, ITGB3, KIAA1919, LOC440563, LOC647020, LRRC69, MAGEF1, MBIP, MCAT, MET, MLL4, MMP3, MTERF, NDUFA10, NEIL1, NLRP7, NOD2, NOX1, NPL, NUP188, OTOA, OXSM, PAFAH2, PARK2, POLQ, POLR3GL, PPEF2, PRMT7, PRSS3, PRSS7, PTCHD3, RBKS, RPS6KC1, SAMD9L, SERPINI2, SGOL2, SIGLEC1, SLAMF6, SLC26A10, SLC4A1AP, SLC6A5, SLC01B3, SMARCD2, SPATA4, SSSX9, STAP2, TAOK1, TIMD4, TMEM56, TNFAIP6, TNFAIP8, TTC21A, USP45, VPS13B, WFDC8, WRN, ZIM2, ZNF311, ZNF451, ZNF491, ZNF582, ZNF599, ZSCAN29; [cell–cell and cell–extracellular matrix adhesion processes]</b>	253 [229]	21858661; 23383274; 28062980; 27913932; 27655433; 25288723; 24470238; 23800003; 27648926; 29363114; 25915596; 25945795; 25050558; 29422015; 22464251; 26576347; 23028338; 25923920; 29325031; 26157685; 25330149; 23409019; 29879995; 29868112; 28202063; 28076423; 24969172; 22527104; 26969729; (73–101) {29}

(Continued on the following page)

**Table 3.** List and counts of identified genes and PubMed IDs by cancer type. (Cont'd)

Cancer type	Authors' identified genes <sup>a</sup> [or pathways]	# genes [unique]	PMID <sup>b</sup> (references) {number of articles}
Colorectal	<i>BRF1</i> ; <b>FANI</b> ; <b>MUTYH</b> ; <i>RPS20</i> ; <i>SETD6</i> ; <b>CENPE</b> ; <i>KIF23</i> ; <b>POLE</b> ; <b>POLD1</b> ; <i>WRN</i> ; <b>ERCC6</b> ; <i>EMR3</i> ; <i>PTPN12</i> ; <i>LRP6</i> ; <i>TTF2/TRIM45/VTCN1/MANIA2</i> ; <b>BRCA2</b> ( <i>FANCD1</i> ); <b>BRIPI</b> ( <i>FANCI</i> ); <b>FANCC</b> ; <i>FANCE</i> ; <i>REV3L</i> ( <i>POLZ</i> ); <b>FANCM</b> ; <b>LAMB4</b> ; <i>LAMC3</i> ; <b>PTCHD3</b> ; <i>TREX2</i> ; <i>IL12RB1</i> ; <i>LIMK2</i> ; <b>POLE2</b> ; <b>POT1</b> ; <i>MRE11</i> ; <i>CDKN1B</i> ; <i>XRCC4</i> ; <i>EPHX1</i> ; <b>NFKBIZ</b> ; <b>SMARCA4</b> ; <b>BARD1</b> ; <i>ACSL5</i> ; <i>ADAMTS4</i> ; <i>ARHGAP12</i> ; <b>ATM</b> ; <i>CYTL1</i> ; <i>DONSON</i> ; <i>INTS5</i> ; <i>MCTP2</i> ; <i>ROSI</i> ; <b>SYNE1</b> ; <i>UACA</i> ; <i>SFXN4</i> ; <i>TWSG1</i> ; <i>PSPH</i> ; <i>NUDT7</i> ; <i>ZNF490</i> ; <i>PRSS37</i> ; <i>CCDC18</i> ; <i>PRADC1</i> ; <i>MRPL3</i> ; <i>AKR1C4</i> ; <i>DDX20</i> ; <i>ZFYVE26</i> ; <i>PIK3R3</i> ; <i>SLC26A8</i> ; <i>ZEB2</i> ; <i>TP53INP1</i> ; <i>SLC11A1</i> ; <i>LRBA</i> ; <i>CEBPZ</i> ; <i>ETAA1</i> ; <i>SEMA3G</i> ; <i>IFRD2</i> ; <i>FAT1</i> ; <b>MLH1</b> ; <b>MUTYH</b> ; <b>MSH6</b> ; <b>MSH2</b> ; <i>LRP5</i> ; <b>ATM</b> ; <i>RYR3</i> ; <i>EIF2AK4</i> ; <i>PRDM1</i> ; <i>RYR2</i> ; <b>BUB1</b> ; <i>RPS6KB2</i> ; <i>DAAMI</i> ; <i>TCF7</i> ; <b>TSC2</b> ; <i>LIG3</i> ; <i>MCC</i> ; <b>MAX</b> ; <b>PARP1</b> ; <i>ETV4</i> ; <i>FZD10</i> ; <i>MAST2</i> ; <i>MTOR</i>	96 [94]	28912018; 26052075; 24691292; 24941021; 28973356; 23637064; 23263490; 26344056; 26901136; 29396139; 27165003; 23585368; 27329137; 25058500; 25749350; 24146633; 29844832; 25892863; (102-119) {18}
Esophageal	<i>KCNJ12/KCNJ18</i> ; <i>GPRIN2</i> ; <b>PTEN</b> ; <i>SMAD7</i> ; <i>DNAH9</i> ; <i>GKAP1</i> ; <i>NFX1</i> ; <i>BAG1</i> ; <i>FUK</i> ; <i>DDOST</i> ; <i>NR4A2</i> ; <i>IL6ST</i> ; <i>FZD3</i> ; <i>SBNO2</i> ; <i>NOTCH3</i> ; <i>STAT3</i> ; <i>CD3D</i> ; <i>MUC16</i> ; <i>LIN28A</i> ; <i>STIL</i> ; <i>ANK1</i> ; <i>CD3E</i> ; <i>PARN</i> ; <i>GDF15</i> ; <i>ITGA6</i> ; <i>LCK</i> ; <i>THBS2</i> ; <i>ALX4</i> ; <b>NOD2</b> ; <i>KALRN</i> ; <i>EGF</i> ; <i>DLK2</i> ; <i>HDAC2</i> ; <i>AKT1</i> ; <i>FBXW7</i> ; <b>MET</b> ; <i>SFRP1</i>	38 [38]	29405996; 25554686; 28165652; 28459198; (120-123) {4}
Gastric	<i>none</i> ; <i>ATP4A</i> ; <i>CTNNA1</i> ; <i>INSR</i> ; <i>FBXO24</i> ; <i>DOT1L</i> ; <b>PALB2</b> ; <b>BRCA1</b> ; <b>RAD51C</b> ; <b>PALB2</b> ; <b>MSH2</b> ; <b>ATR</b> ; <b>NBN</b> ; <b>RECQL5</b> ; <i>DZIP1L</i> ; <i>PCOLCE2</i> ; <i>TAS2R7</i> ; <i>TRIOBP</i> ; <i>IGSF10</i> ; <b>NOTCH1</b> ; <i>SF3A1</i> ; <i>GAL3ST1</i> ; <b>NOTCH1</b> ; <i>OR13C8</i> ; <i>EPB41L4B</i> ; <i>SEC16A</i>	25 [23]	28875981; 25678551; 23208944; 25576241; 28024868; 29706558; 26872740; (124-130) {7}
Glands	<b>MAX</b> ; <i>HOXA11</i> ; <i>NCOA4</i> ; <i>PCSK7</i> ; <i>CENPQ</i> ; <i>GALC</i> ; <i>GOLGA5</i> ; <i>GUCA1C</i> ; <i>RHCE</i> ; <i>DDX12</i> ; <i>FAM22F</i> ; <i>FAM71F1</i> ; <i>MDN1</i> ; <b>MLL3</b> ; <b>SYNE1</b> ; <i>IGSF3</i> ; <i>STXBP5L</i> ; <i>PCDH12</i> ; <i>HAVCR1</i> ; <i>SMPD2</i> ; <i>OPRM1</i> ; <i>GTF2IRD2/GTF2IRD2B</i> ; <i>TAS2R43</i> ; <i>KMT5A</i> ; <b>NCOR2</b> ; <i>BTBD6</i> ; <i>PHBDF1</i> ; <b>TSC2</b> ; <i>RAMP2</i> ; <i>CST2</i> ; <i>SHANK3</i>	32 [32]	21685915; 23707928; 28210977; (131-133) {3}
Head and neck	<b>MLL3</b> ; <b>MSTIR</b> ; <i>EXO1</i> ; <i>HLTF</i> ; <i>TDP2</i> ; <b>RAD52</b> ; <i>PER1</i> ; <b>MSH2</b> ; <i>ERCC5</i> ; <b>MSH6</b> ; <i>RAD17</i> ; <b>PMS2</b> ; <b>NBN</b> ; <i>DCLRE1C</i> ; <b>ALKBH3</b> ; <b>LIG4</b> ; <i>XRCC3</i> ; <b>FANCM</b> ; <b>PALB2</b> ; <b>BRCA1</b> ; <b>BRIPI</b> ; <b>TP53</b> ; <i>XRCC1</i> ; <b>CHEK2</b> ; <i>FAAP20</i> ; <i>WDR48</i> ; <b>FANCD2</b> ; <b>FANCA</b> ; <i>SHPRH</i> ; <b>MMS19</b> ; <b>ALKBH3</b> ; <b>PARP4</b> ; <b>RECQL5</b> ; <i>NUDT1</i> ; <b>PARP2</b> ; <b>POLE2</b>	36 [35]	26014803; 26951679; 29747023; (134-136) {3}
Intestinal	<i>IPMK</i> ; <b>MUTYH</b>	2 [2]	25865046; 28634180; (137, 138) {2}
Kidney	<b>BAP1</b> ; <i>CTR9</i> ; <i>REST</i> ; <b>DICER1</b>	4 [4]	23684012; 25099282; 26551668; 26566882; (139-142) {4}
Liver	<b>DICER1</b>	1 [1]	28012864; (143) {1}
Lung	<b>CHEK2</b> ; <i>HER2</i> ; <i>IBSP</i> ; <b>MET</b> ; <b>PARK2</b> ; <i>YAP1</i> ; <i>PROM1</i> ; <i>CRTC2</i> ; <i>CACNB2</i> ; <b>CENPE</b> ; <i>LCT</i> ; <i>MAST1</i> ; <i>ARHGEF5</i> ; <i>ANKRD20A2</i> ; <i>ZNF595</i> ; <i>ZNF812</i> ; <b>MYO18B</b> ; <i>BPIFB1</i> ; <i>CHD4</i> ; <b>PARP1</b> ; <i>NUDT1</i> ; <b>RAD52</b> ; <i>MF12</i> ; <b>FANCL</b> ; <b>FANCG</b> ; <b>FANCC</b> ; <b>FANCF</b> ; <b>BRCA2</b> ; <b>PALB2</b> ; <b>FANCA</b> ; <b>BRCA1</b> ; <b>RAD51C</b> ; <b>BRIPI</b> ; <i>BAG6</i> ; <i>SPEN</i> ; <i>WISP3</i> ; <b>JAK2</b> ; <i>TCEB3C</i> ; <i>NELFE</i> ; <i>TAF1B</i> ; <i>EBLN2</i> ; <i>GON4L</i> ; <i>NOP58</i> ; <i>RBMX</i> ; <i>KIAA2018</i> ; <b>ZNF311</b> ; <b>TP53</b> ; <b>BRCA1</b> ; <b>BRCA2</b> ; <i>ERCC4</i> ; <i>EXT1</i> ; <i>HNF1A</i> ; <b>PTCH1</b> ; <i>SMARCB1</i> ; <i>ABCC10</i> ; <i>ATP7B</i> ; <i>CACNA1S</i> ; <b>CFTR</b> ; <i>CLIP4</i> ; <i>COL6A1</i> ; <i>COL6A6</i> ; <i>GCN1</i> ; <i>GJB6</i> ; <i>RYR1</i> ; <i>SCN7A</i> ; <i>SEC24A</i> ; <i>SPI00</i> ; <i>TIN</i> ; <i>USH2A</i>	69 [67]	27900359; 24317180; 26717996; 28294470; 25640678; 26056182; 24484648; 26178433; 29054765; 29667179; 30425093; 24954872; 30032850; (144-156) {13}
Melanoma	<i>none</i> ; <b>BAP1</b> ; <i>GOLM1</i> ; <b>POLE</b> ; <b>POT1</b> ; <b>POT1</b> ; <i>RAD51B</i> ; <i>TP53AIP1</i> ; <i>VPS41</i> ; <b>CDKN2A</b> ; <b>BAP1</b> ; <i>EBF3</i> ; <i>MC1R</i> ; <b>MITF</b> ; <b>BRCA2</b> ; <i>MTAP</i>	15 [13]	29764119; 23977234; 29659923; 26251183; 24686846; 24686849; 25600502; 29359367; 25303718; 29522175; 29317335; (157-167) {11}
Multiple familial <sup>c</sup>	<b>ERCC3</b> ; <b>BAP1</b> ; <b>POT1</b> ; <b>MLL3</b> ; <b>POLE</b> ; <b>POLE</b> ; <i>HNRNPAO</i> ; <i>WIFI</i>	8 [7]	28911001; 22889334; 28389767; 23429989; 24788313; 25860647; 25716654; (168-174) {7}
Multiple types <sup>d</sup>	<b>BRCA1</b> ; <b>BRCA2</b> ; <b>ATM</b> ; <b>BRIPI</b> ; <b>PALB2</b> ; <i>CNKSR1</i> ; <i>EME2</i> ; <i>MRE11A</i> ; <b>MSH6</b> ; <b>PIK3C2G</b> ; <b>RAD51C</b> ; <b>RAD51D</b> ; <b>XRCC2</b> ; <b>APC</b> ; <b>ATM</b> ; <b>BAP1</b> ; <b>BRCA1</b> ; <b>BRCA2</b> ; <b>BUB1B</b> ; <i>FH</i> ; <i>HFE</i> ; <b>MAX</b> ; <b>MET</b> ; <b>MSH6</b> ; <b>NF1</b> ; <b>PALB2</b> ; <b>PTEN</b> ; <b>RET</b> ; <b>SDHA</b> ; <b>SDHB</b> ; <b>SDHD</b> ; <b>TP53</b> ; <b>VHL</b> ; <b>WRN</b>	34 [29]	26689913; 29625052; (175, 176) {2}
Pediatric	<b>APC</b> ; <b>DICER1</b> ; <b>TP53</b> ; <b>BRCA2</b> ; <b>CHEK2</b> ; <b>BAP1</b> ; <b>BUB1</b> ; <i>ETV6</i> ; <i>ACTB</i> ; <i>ARID1A</i> ; <i>EP300</i> ; <i>EZH2</i> ; <i>APOB</i> ; <i>SCN2A</i> ; <i>SPRED1</i> ; <i>KDM3B</i> ; <b>TYK2</b> ; <b>TP53</b> ; <b>APC</b> ; <b>BRCA2</b> ; <b>NF1</b> ; <b>PMS2</b> ; <b>RB1</b> ; <i>RUNX1</i> ; <i>ALK</i> ; <b>BRCA1</b> ; <i>CDH1</i> ; <i>KRAS</i> ; <b>MSH2</b> ; <b>MSH6</b> ; <i>NF2</i> ; <i>NRAS</i> ; <b>PALB2</b> ; <b>PTCH1</b> ; <b>RET</b> ; <b>SDHA</b> ; <b>SDHB</b> ; <b>VHL</b>	39 [36]	29351919; 26580448; (177, 178) {2}
Ovarian	<b>FANCM</b> ; <b>SMARCA4</b> ; <b>RAD51D</b> ; <b>ATM</b> ; <b>FANCM</b> ; <b>BRCA1</b> ; <b>TP53</b> ; <b>NF1</b> ; <i>MAP3K4</i> ; <i>CDKN2B</i> ; <b>MLL3</b> and other ~220 genes not stated	~230	28881617; 24658002; 28591191; 24448499; (179-182) {4}

(Continued on the following column)

Downloaded from <http://aacrjournals.org/cebp/article-pdf/29/8/1519/1945786/1519.pdf> by guest on 22 June 2024

**Table 3.** List and counts of identified genes and PubMed IDs by cancer type. (Cont'd)

Cancer type	Authors' identified genes <sup>a</sup> [or pathways]	# genes [unique]	PMID <sup>b</sup> (references) {number of articles}
Ovarian and breast	<i>ATM</i> , <i>MYC</i> , <i>PLAU</i> , <i>RAD1</i> , <i>RRM2B</i> ; <i>ATM</i> , <i>PALB2</i> , <i>CHEK2</i> , <i>MSH6</i> , <i>TP53</i> , <i>RAD51C</i>	11 [10]	27782108; 30128536; (183, 184) {2}
Pancreatic	<i>ATM</i> ; <i>BRCA2</i> ; <i>ATM</i> , <i>PALB2</i> ; <i>FAN1</i> , <i>NEK1</i> , <i>RHNO1</i> ; <i>ATM</i> , <i>CDKN2A</i> , <i>PALB2</i> , <i>BRCA2</i> , <i>BUB1B</i> , <i>CPA1</i> , <i>FANCC</i> , <i>FANCG</i> ; <i>ATM</i> , <i>BRCA1</i> , <i>BRCA2</i> , <i>CFTR</i> , <i>MSH2</i> , <i>MSH6</i> , <i>PALB2</i> , <i>PMS2</i> , <i>TP53</i>	24 [17]	22585167; 29074453; 23561644; 26546047; 26658419; 27449771; (185-190) {6}
Prostate	<i>none</i> ; <i>BTNL2</i> ; <i>BRCA2</i> , <i>ATM</i> , <i>NBN</i> ; <i>BRCA2</i> , <i>HOXB13</i> , <i>TRRAP</i> , <i>ATP1A1</i> , <i>BRIPI</i> , <i>FANCA</i> , <i>FGFR3</i> , <i>FLT3</i> , <i>HOXD11</i> , <i>MUTYH</i> , <i>PDGFRA</i> , <i>SMARCA4</i> , <i>TCF3</i> ; <i>HOXB13</i> , <i>TANGO2</i> , <i>OR5H14</i> , <i>CHAD</i> ; <i>LRCC46</i> , <i>PARP2</i> , <i>BLM</i> , <i>KIF2B</i> , <i>CYP3A43</i> ; <i>ATM</i> , <i>ATR</i> , <i>BRCA2</i> , <i>FANCL</i> , <i>MSR1</i> , <i>MUTYH</i> , <i>RBI</i> , <i>TSHR</i> , <i>WRN</i> ; <i>CYP3A43</i> , <i>HEATR5B</i> , <i>GPR124</i> , <i>HKR1</i> , <i>PARP2</i> , <i>PCTP</i> , <i>MCRS1</i> , <i>DOK3</i> , <i>ATRIP</i> , <i>PLEKHH3</i> ; <i>TE2</i> , <i>CEP63</i> , <i>ELK4</i> , <i>NUBP2</i> , <i>SPRR3</i> , <i>TRBV7-7</i> , <i>BRCA1</i> , <i>LAMB3</i> , <i>MSH6</i> , <i>PARP2</i> , <i>DLEC1</i> , <i>ZSWIM2</i> ; <i>ACACA</i> , <i>AKR1C1</i> , <i>ALG13</i> , <i>ALG6</i> , <i>APTX</i> , <i>ASXL1</i> , <i>ATP6VOA2</i> , <i>B3GAT3</i> , <i>BGLAP</i> , <i>BLM</i> , <i>BRCA1</i> , <i>CRISP3</i> , <i>CTBP1</i> , <i>CYP1B1</i> , <i>DDDB2</i> , <i>DOLK</i> , <i>EFCAB6</i> , <i>ELAC2</i> , <i>FANCA</i> , <i>FANCL</i> , <i>FBXW7</i> , <i>FLT3</i> , <i>GSTA1</i> , <i>HDAC9</i> , <i>HSD3B1</i> , <i>IDE</i> , <i>KDR</i> , <i>MBD5</i> , <i>MGAT2</i> , <i>MSH3</i> , <i>MYH14</i> , <i>NCOR2</i> , <i>NEIL3</i> , <i>NOTCH2</i> , <i>NRIP1</i> , <i>PALB2</i> , <i>PAPSS2</i> , <i>PIAS3</i> , <i>RAD51D</i> , <i>RAD54L2</i> , <i>RNASEL</i> , <i>SCN11A</i> , <i>SP1</i> , <i>SULT1E1</i> , <i>TGIF1</i> , <i>TLN1</i> , <i>TP53BP1</i> , <i>TSC1</i> , <i>BE2D3</i> , <i>UBE2V2</i>	107 [95]	26604137; 23833122; 29915322; 27701467; 27902461; 25111073; 27084275; 26585945; 27486019; 26485759; (191-200) {10}
Sarcoma <sup>e</sup>	<i>POT1</i> ; <i>SMARCA4</i> ; <i>STAT4</i> ; <i>CDKN2A</i> , <i>PDGFRA</i> ; <i>APC</i> , <i>BLM</i> , <i>BRCA1</i> , <i>BRIPI</i> , <i>ERCC3</i> , <i>EXT2</i> , <i>FANCC</i> , <i>FANCD2</i> , <i>FANCM</i> , <i>FLCN</i> , <i>MITF</i> , <i>PMS2</i> , <i>POLE</i> , <i>PTCH2</i> , <i>PTPN11</i> , <i>RAD51</i> , <i>RAD51D</i> , <i>RET</i> , <i>SLX4</i> , <i>TINF2</i> , <i>TP53</i> , <i>WRAP53</i>	27 [27]	26403419; 23775540; 25492914; 28592523; 28125078; (201-205) {5}
Testicular	<i>none</i> ; <i>PKD1</i> ; <i>PLEC</i> , <i>DNAH7</i> , <i>EXO5</i> ; <i>DNAAF1</i> , <i>LRR6C</i> , <i>CNTRL</i> , <i>DRC1</i> , <i>DYNC2H1</i> , <i>CEP290</i> , <i>MAP4</i>	11 [11]	29433971; 27577987; 29761480; 27996046; (206-209) {4}
Thyroid	<i>C14orf93</i> ( <i>RTFC</i> ); <i>HABP2</i> ; <i>MAP2K5</i> ; <i>MET</i> ; <i>PARP4</i> ; <i>SRRM2</i> ; <i>USF3</i> ; [extracellular matrix organization and DNA repair]	7 [7]	27864143; 26222560; 30132833; 29219214; 26699384; 26135620; 28011713; 28402931; (210-217) {8}

<sup>a</sup>The definition of "identified" is based on the authors' choice and emphasis in presenting their results and not on an evaluation of the quality of the results. Bold gene symbols indicate genes reported in more than one article, black if only within cancer type, blue if both within and across cancer types, and red if only across cancer types.

<sup>b</sup>Articles are referenced in the same order as the results separated by semicolons.

<sup>c</sup>Single families with multiple cancer types: 28911001 [adrenocortical, breast (Li-Fraumeni syndrome)]; 22889334 (melanoma, paraganglioma); 28389767 (melanoma, thyroid, breast); 23429989 (colon, blood); 24788313 (colon, ovary, endometrium, brain); 25860647 (colon, ovary, pancreas, small intestine); and 25716654 (prostate, breast, colon, pancreas, melanoma).

<sup>d</sup>Large studies of the genetic landscape across cancer types: 26689913 (across 12 cancer types) and 29625052 (across 33 cancer types).

<sup>e</sup>Sarcoma types: soft tissue (28592523), Ewing (28125078), cardiac (26403419), Kaposi (25492914), and rhabdoid (23775540).

stated that they were not able to identify variants or genes in the studied cases, and the remaining two (1%) articles pointed to molecular or functional pathways of possible relevance to the studied cancer type. The 176 articles indicated as primary findings (Supplementary Table S3) a total of approximately 2,095 variants (average 11, median 3, range 1–222 per article) and approximately 1,215 (~954 unique) genes (average 6, median 1, range 1–<222 per article). An exact count of variants and genes identified was not feasible due to incomplete counts and/or variant nomenclature in some of the articles. For the 99 articles that studied more than one high-risk family, and reported the information, the identified variants/genes accounted on average for 25% of the families evaluated in discovery and replication phases combined (we excluded from this analysis 43 articles that studied only a single family and did not attempt replication, Supplementary Fig. S10). Regarding the prevalence of the identified variants among controls, 27 (16% of 176) articles did not sequence these variants in controls, 50 (28%) sequenced only some unaffected relative of the cases, 37 (21%) did not report how many of the sequenced controls carried the investigated variants; in the remaining articles the variants'

frequency in unrelated controls averaged 0.015 with median 0 (Supplementary Table S3).

Overall, 106 genes were identified in two or more articles (see bolded gene symbols in Table 3). The five genes reported by more than 10 articles are well-established cancer susceptibility genes (i.e., *ATM*, *BRCA2*, *BRCA1*, *TP53*, and *PALB2* were observed in 12%, 12%, 9%, 9%, and 8% of the articles, respectively). When the analysis was restricted to the articles that used a fully agnostic, not candidate, analytic approach, these genes were observed less frequently (6%, 4%, 7%, 0%, 2%, respectively) and other less established genes were more frequently observed (>4%; i.e., *PMS2*, *IGSF22*, *ABCA10*, *ACAN*, and *PABPC3*). We also observed 43 variants in 22 genes that were independently identified in two or three articles (Table 4). While some of the observed pleiotropic effects are well established (e.g., *PALB2* and *BRCA1/2* for breast, ovarian, prostate, and pancreatic cancer), others are potentially novel, such as *BRCA2* for melanoma and head and neck cancers, *MUTYH* for prostate and small intestine cancers, and *KDR* for prostate cancer and Hodgkin lymphoma.



**Table 4.** Germline variants identified in two or more of the 186 reviewed articles.

Gene	Variant	Allele frequency	Articles' PMID	Cancer type
<i>PALB2</i>	chr16, c.172_175delTTGT, p.Q60fs, rs1214293842	0.000042	25330149, 30128536	Breast, Ovarian and breast
<i>PALB2</i>	chr16, c.509_510delGA, p.R170fs, rs863224790	0.000037	25330149, 30128536	Breast, Ovarian and breast
<i>PALB2</i>	chr16, c.1240C>T, p.R414*, rs180177100	0.000009	27449771, 30128536	Pancreatic, Ovarian and breast
<i>PALB2</i>	chr16, c.3256C>T, p.R1086*, rs587776527	0.000009	23561644, 30128536	Pancreatic, Ovarian and breast
<i>PALB2</i>	chr16, c.3004_3007delGAAA, p.E1002Tfs	0	26485759, 30128536	Prostate, Ovarian and breast
<i>PALB2</i>	chr16, c.3549C>A, p.Y1183*, rs118203998	0.000009	26689913, 30128536	Multiple_12, Ovarian and breast
<i>PALB2</i>	chr16, c.424A>T, p.K142*	0	26689913, 30128536	Multiple_12, Ovarian and breast
<i>BRCA2</i>	chr13, c.658_659delGT, p.V220fs, rs876660049	0.000028	26689913, 28202063	Multiple_12, Breast
<i>BRCA2</i>	chr13, c.6275_6276delTT, p.Leu2092fs, rs11571658	0.000065	25330149, 29915322	Breast, Prostate
<i>BRCA2</i>	chr13, c.9246_9247insA, p.T3085fs, rs80359752	0	25330149, 29915322	Breast, Prostate
<i>BRCA2</i>	chr13, c.865A>C, p.N289H, rs766173	0.052597	29317335, 29747023	Melanoma, Head and neck
<i>BRCA2</i>	chr13, c.9294C>G, p.Y3098*, rs80359200	0.000009	26580448, 29625052	Pediatric, Multiple_33
<i>ATM</i>	chr11, c.5071A>C, p.S1691R, rs1800059	0.001937	28202063, 28652578	Breast, Blood (CLL)
<i>ATM</i>	chr11, c.170G>GA, p.W57*	0	22585167, 30128536	Pancreatic, Ovarian and breast
<i>ATM</i>	chr11, c.6095G>GA, p.R2032K, rs139770721	0.000027	22585167, 29625052	Pancreatic, Multiple_33
<i>ATM</i>	chr11, c.6100C>T, p.R2034*, rs532480170	0.000009	27913932, 30128536	Breast, Ovarian and breast
<i>ATM</i>	chr11, g.108155008_delG, p.E1267fs	0	28652578, 29625052	Blood (CLL), Multiple_33
<i>BRCA1</i>	chr17, c.1067A>G, p.Q356R, rs1799950	0.045162	25923920, 26485759	Breast, Prostate
<i>BRCA1</i>	chr17, c.4065_4068delTCAA, p.N1355fs, rs886040195	0.000018	26689913, 29625052	Multiple_12, Multiple_33
<i>BRCA1</i>	chr17, c.1054G>T, p.E352*, rs80357472	0.000009	26689913, 29625052	Multiple_12, Multiple_33
<i>BRCA1</i>	chr17, c.68_69delAG, p.E23fs, rs80357410	0.000175	26689913, 29625052	Multiple_12, Multiple_33
<i>TP53</i>	chr17, c.733C>T, p.G245S, rs28934575	0	26580448, 29351919, 29602769	Pediatric, Pediatric, Brain
<i>TP53</i>	chr17, c.524G>A, p.R175H, rs28934578	0	26580448, 30128536	Pediatric, Ovarian and breast
<i>TP53</i>	chr17, c.743G>A, p.R248Q, rs11540652	0.000009	26580448, 30128536	Pediatric, Ovarian and breast
<i>FANCM</i>	chr14, c.5101C>T, p.Q1701*, rs147021911	0.001530	25288723, 28881617	Breast, Ovarian
<i>FANCM</i>	chr14, c.5791C>T, p.R1931*, rs144567652	0.001181	28591191, 28881617	Ovarian, Ovarian
<i>KAT6B</i>	chr10, c.4546G>T, p.D1516Y	0	23800003, 24969172	Breast, Breast
<i>KAT6B</i>	chr10, c.4729C>T, p.R1577C	0	23800003, 24969172	Breast, Breast
<i>POT1</i>	chr7, c.1851_1852delTAA, p.D617fs, rs758673417	0.000009	25482530, 27329137	Glioma, Colorectal
<i>MSH6</i>	chr2, c.3261delC, p.F1088fs	0	26689913, 30128536	Multiple_12, Ovarian and breast
<i>CHEK2</i>	chr22, c.1100delC, p.T367fs	0	22527104, 29351919, 30128536	Breast, Pediatric, Breast/ovarian
<i>RAD51D</i>	chr17, g.33433425G>A, p.R206*, rs387906843	0.000017	26689913, 28591191	Multiple_12, Ovarian
<i>FANCC</i>	chr9, c.C553C>T, p.R185*, rs121917783	0.000064	23028338, 28125078	Breast, Sarcoma (Ewing)
<i>MUTYH</i>	chr1, c.1187G>A, p.G396D	0	27084275, 28634180	Prostate, Intestine (small)
<i>BLM</i>	chr15, c.1933C>T, p.Q645*, rs373525781	0.000018	23028338, 28125078	Breast, Sarcoma (Ewing)
<i>TYK2</i>	chr19, c.2279C>T, p.P760L	0	27733777, 29351919	Blood (ALL), Pediatric
<i>MAX</i>	chr14, c.223C>T, p.R75*	0	21685915, 29625052	Pheochromocytoma, Multiple_33
<i>NOTCH2</i>	chr1, c.3625T>G, p.F1209V, rs147223770	0.003217	26485759, 29868112	Prostate, Breast
<i>XRCC2</i>	chr7, c.96delT, p.F32fs, rs774296079	0.000075	25330149, 26689913	Breast, Multiple_12
<i>RET</i>	chr10, c.2370G>C, p.L790F, rs75030001	0.000009	26580448, 28125078	Pediatric, Sarcoma (Ewing)
<i>GPRC5A</i>	chr12, c.183delG, p.R61fs, rs527915306	0.002113	22527104, 24470238	Breast, Breast
<i>KDR</i>	chr4, c.3193G>A, p.A1065T, rs56302315	0.000324	26485759, 27365461	Prostate, Blood (HL)
<i>MITF</i>	chr3, c.952G>A, p.E318K, rs149617956	0.001330	28125078, 29317335	Sarcoma (Ewing), Melanoma

Abbreviations: ALL, acute lymphoblastic leukemia; CLL, chronic lymphocytic leukemia; HL, Hodgkin lymphoma.

## Discussion

### Methodologic variability across reviewed articles

One major observation from this review was that the criteria used to identify variants and genes presented by the authors as having a role in cancer susceptibility varied dramatically across studies (Fig. 1). In addition, most reviewed articles lacked sensitivity analyses to assess the variability in results by changes in variant/gene selection strategy or a justification for the choice of criteria and thresholds adopted. For example, although restricting the analysis to rare variants is justified in principle by the fact that high-risk/high-penetrance variants are very rare in the general population, no clear justification (e.g., based on disease penetrance estimates) was usually given for the exact choice of MAF thresholds, which can impact both false positives and negatives. In addition, requiring that the selected variants be completely absent from internal or publicly available control datasets (25% of articles)

may also lead to false negatives, given that known disease related variants are observed in these datasets. Moreover, differences in methodologies, such as, study design, sequencing technologies, depth of coverage, human genome reference used, annotation software, variant calling methods, *in silico* prediction tools, have been reported to lead to differences in findings (218–219). Similarly, although the choice of transcript set and annotation software have been quantified to have a substantial effect on variant annotation and impact on the analysis of genome sequencing studies (220), none of the reviewed articles examined or discussed these potential effects. Although this literature review spanned a decade wherein underlying technologies, costs, and bioinformatic pipelines evolved significantly, we note that 74% of the reviewed articles were published in the years 2015 to 2018. When restricting the analyses to this group of recent articles, we observed similar results.

Downloaded from <http://aacrjournals.org/cebp/article-pdf/29/8/1519/1945786/1519.pdf> by guest on 22 June 2024

### Importance of developing consensus on standards

The observed wide variation and inconsistencies in approaches and strategies underscore the importance of establishing a consensus on standards for filtering strategies and rationale for variant identification (e.g., justification for the criteria and thresholds used). The disadvantage of using different methodologies to identify germline susceptibility genes is that it limits the ability to compare results across studies. While initiatives by the American College of Medical Genetics and Genomics (221) and the NIH (222) have developed standards to assign a pathogenicity status to a given variant based on the available literature and annotations, to our knowledge, there has not been an attempt to set standards or a framework for agnostic searches of susceptibility variants or genes. On the basis of the present systematic review, we would recommend that articles in this field: (i) report information for all the relevant components described in **Table 1** and Supplementary Table S1; (ii) include a complete list of identified variants/genes and a count of the individuals carrying those in a format similar to Supplementary Table S3; (iii) report and explain the choice of variants/genes filtering criteria and thresholds, including sensitivity analysis when warranted.

### Variants and genes identified in the reviewed articles

Approximately 95% of the reviewed studies reported identifying susceptibility variants or genes in the studied cancer cases. However, this observation may reflect general publication bias. Overall, about 2,000 variants and 1,000 unique genes were reported as primary findings by the authors. Breast cancer studies reported the highest number of genes, possibly reflective of the large proportion of published studies rather than the underlying genetic architecture. Notably, one hundred genes were found in more than one article (see bolded gene symbols in **Table 3**), indicating that results are recurrent within each cancer type and suggestive of pleiotropic effects across cancer types. Some of these observations may also be due to chance and/or to the wide adoption of variant/gene selection approaches based on known candidate variants, genes, or pathways. Indeed, when restricting the analysis to the articles that used a fully agnostic analysis approach, we observed a decrease in the relative frequency of reporting of these genes and an increase in relative frequency for less established genes. This observation may illustrate that more novel genes could be discovered by using a more expansive analysis approach. In addition, we found that 43 variants (**Table 4**) were each identified in two or more articles across cancer types. The identified variants/genes accounted on average for 25% of the families evaluated in both discovery and replication, suggesting that the fraction of families explained by the genes identified through exome/genome-wide sequencing may have increased since previous linkage analysis and candidate gene sequencing results (10%–25% of families depending on the cancer type; ref. 223). The results collectively show that important progress has been made in the identification of cancer susceptibility genes and that pleiotropy is a common phenomenon in genetic cancer susceptibility. Nevertheless, the progress made to date is not without caveats.

### Challenges limiting progress in variant/gene identification

This review reveals scientific gaps and challenges in the body of literature. Of note, many (especially rare) cancer types remain understudied (or under published) and over 75% of cancer-prone families remain unexplained. While the lack of identification of mutations for cancer in heavily loaded families could reflect a polygenic or omnigenic architecture in these families, several additional challenges may have limited further progress in identifying germline variants associated with cancer, as indicated by the suspected publication bias (95%

articles reported positive findings) and by the limited number of articles identified (only 186 articles across 10 years and all cancer types). First, through careful review of the literature, we observed variation in study design and case selection, even within studies. For example, many of the studies included in this review used familial cases in the discovery phase before switching to unselected cases in the replication phase (perhaps due to a lack of additional familial samples or funds), which may introduce etiologic heterogeneity (e.g., familial cases may carry different and/or more penetrant variants/genes than unselected cases) and may, in part, explain the lack of replication for some of the reviewed studies. In addition, including suitable control populations is important to ascertain magnitude of risk, whereas the frequency of the identified variants in such controls was reported only in one third of the articles. Second, focusing exclusively on the exome (only 10 of the articles were WGS) may be a limitation in complex trait genetics for which noncoding genetic variation is believed to play a larger role than in Mendelian genetics (224, 225)—a hypothesis that still needs to be verified for rare variants specifically. A third aspect that may have limited progress is the widespread use of candidate analysis approaches that focus the discovery analysis on known variants or genes or pathways by leveraging relevant existing information to select the resulting variants/genes. Although articles that used a candidate sequencing approach were excluded, the use of candidate analysis approaches was reported in 65% of the reviewed articles. These challenges (lack of genome-wide and agnostic studies) may be due to the fact that researchers do not yet have the tools to examine agnostically the whole exome or genome effectively. Alternatively, the researchers may have had specific reasons to focus on candidate regions of interest. Whatever the origin, an important consequence should be acknowledged: much of the human exome (and genome) remains unexplored or untested for cancer. Finally, authors frequently stated a need for additional research to replicate their findings in larger and more homogeneous (e.g., by race/ethnicity or cancer histology) study populations. Indeed, although the majority of the reviewed articles used in the discovery phase a familial study design (that does not require cancer case numbers as large as unselected case-control study designs), 39% of articles exome/genome sequenced only a single family and 26% only a single member per family. Increasing the number of sequenced families and of cancer cases within each family may be an important avenue for future studies.

### Technical considerations

From a technical point of view, the differences in utilization of the various technologies are dependent on the timing of their development and subsequent replacement by the next capture kit or sequencer version. We showed that the reviewed articles reported a technical validation rate of about 80% for studies that tested over fifty variants versus over 90% for studies that tested fewer variants; the difference may be due to prevalidation manual inspection steps (e.g., through IGV) being more feasibly applied to a limited number of variants. This observation suggests the importance for researchers using current sequencing technologies at a genome-wide scale to technically validate any observed variants. Another technical limitation stems from within study aggregation of samples across multiple sequencing experiments, as this approach can generate biases in variant detection and false positives/negatives in variant-cancer associations, particularly for WES datasets that vary also in capture efficiency. Several strategies to control for biological and technical heterogeneity and to minimize calling discordance and erroneous findings were described in the reviewed articles, including checking for comparable depths and rare variants detection and performing alignment and variant

calling of all samples simultaneously. One notable point, reinforced by the observed lack of data sharing for over 80% of the reviewed articles, is the importance of saving, storing and being able to access BAM or CRAM files of the studied datasets for the wider research community, both for publicly and privately sponsored datasets. Advancements in long-range sequencing and other new technologies may also help address some of the described technical shortcomings in the future and influence near-term approaches to genomic analyses (226).

#### Importance of functional validation

Finally, most reviewed articles acknowledged the importance of functional validation (e.g., through *in vitro* and *in vivo* models) to determine whether the function of the mutated gene product is consistent with the cancer of interest and to inform the interpretation of the reported findings. Even though 65% of the articles did attempt some type of experimental validation for the final selection of variants/genes, the reported functional results were usually not considered definitive by the authors. In fact, most articles described the need for additional functional studies to determine whether the identified genes or variants play a causal role in carcinogenesis and to describe the mechanisms for these variants to impact disease.

#### Limitation and strengths of this literature review

Limitations of our literature review include: the lack of access to primary data, and consequent inability to systematically evaluate how different filtering choices would lead to different results in these studies; likely publication bias toward non-null results; use of the authors definition of “identified gene or variant” that varied greatly across the reviewed articles; the exclusion of articles in which only a single cancer case was sequenced, which although usually case reports, can also in principle lead to the identification of novel cancer susceptibility genes [e.g., *PALB2* (227), or *NPAT* (228)]. Strengths of our

review are the systematic inclusion/exclusion approach, the comprehensive key term search, and thorough data abstraction.

## Conclusions

In conclusion, the findings from this review indicate a growth in usage of NGS technologies at the exome/genome scale to identify genes associated with cancer risk. Nevertheless, progress has been limited by a range of challenges inherent in the field. The review highlights several important next steps including establishing consensus on standards for use and reporting of filtering strategies, describing rationale for variant identification, developing analytic methods that truly mine the whole exome/genome, improving the accuracy and cross-studies interoperability of current sequencing technologies, sharing of the primary data with the research community, and performing extensive variant functional validation. It also points to the untapped potential in conducting studies with more/larger families and in more diverse populations and cancers types, harmonizing results across studies, and expanding searches beyond a candidate analysis approach.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Acknowledgments

This research was supported by the NCI, NIH at the Division of Cancer Control and Populations Sciences, the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, and the Scientific Consulting Group (to N.I. Simonds; contract number HHSN2612014000111).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received December 14, 2019; revised March 17, 2020; accepted May 13, 2020; published first May 28, 2020.

## References

- Meztker ML. Sequencing technologies—the next generation. *Nat Rev Genet* 2010;11:31–46.
- Levy SE, Myers RM. Advancements in next-generation sequencing. *Annu Rev Genom Hum Genet* 2016;17:95–115.
- Bertier G, Hetu M, Joly Y. Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views. *BMC Med Genomics* 2016;9:52.
- Stadler ZK, Schrader KA, Vijai J, Robson ME, Offit K. Cancer genomics and inherited risk. *J Clin Oncol* 2014;32:687–98.
- Alexandrov LB, Stratton MR. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev* 2014;24:52–60.
- Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. *J Hum Genet* 2012;10:621–32.
- Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 2012;20:490–7.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500:415–21.
- Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* 2014;15:585–98.
- Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial risk and heritability of cancer among twins in nordic countries. *JAMA* 2016;315:68–76.
- Susswein LR, Marshall ML, Nusbaum R, Vogel Postula KJ, Weissman SM, Yackowski L, et al. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet Med* 2016;18:823–32.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747–53.
- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *PLoS Med* 2009;6:e1000097.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009, 25:1754–60.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Res* 2010;38:e164.
- The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;526:68–74.
- Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 2012;493:216–20.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T; the Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285.

21. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. *Nat Biotechnol* 2011;29:24–26.
22. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012;40:W542–7.
23. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–49.
24. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014;11:361–2.
25. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5.
26. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Research* 2012;40:D930–4.
27. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research* 2012;22:1790–7.
28. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: protein structure and function prediction. *Nat Method* 2015;12:7–8.
29. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics* 2016;54:5.6.1–37.
30. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014;42:D980–5.
31. McKusick VA. Mendelian inheritance in man: a catalog of human genes and genetic disorders. 12th ed. Baltimore: Johns Hopkins University Press; 1998.
32. Wardell CP, Fujita M, Yamada T, Simbolo M, Fassan M, Karlic R, et al. Genomic characterization of biliary tract cancers identified driver genes and predisposing mutations. *J Hepatol* 2018;68:959–69.
33. Chang VY, Basso G, Sakamoto KM, Nelson SF. Identification of somatic and germline mutations using whole exome sequencing of congenital acute lymphoblastic leukemia. *BMC Cancer* 2013;13:55.
34. Sabri S, Keyhani M, Akbari MT. Whole exome sequencing of chronic myeloid leukemia patients. *Iran J Public Health* 2016;45:346–52.
35. Ristolainen H, Kilpivaara O, Kamper P, Taskinen M, Sarrinen S, Leppä S, et al. Identification of homozygous deletion in ACAN and other candidate variants in familial classical Hodgkin lymphoma by exome sequencing. *Br J Haematol* 2015;170:428–31.
36. Tiao G, Improgo MR, Kasar S, Poh W, Kamburov A, Landau DA, et al. Rare germline variants in ATM are associated with chronic lymphocytic leukemia. *Leukemia* 2017;31:2244–7.
37. Pathak A, Pemov A, McMaster ML, Dewan R, Ravichandran S, Pak E, et al. Juvenile myelomonocytic leukemia due to a germline CBL Y371C mutation: 35-year follow-up of a large family. *Hum Genet* 2015;137:775–87.
38. Pathak A, Seipel K, Pemov A, Dewan R, Brown C, Ravichandran S, et al. Whole exome sequencing reveals a C-terminal germline variant in CEBPA-associated acute myeloid leukemia: 45-year follow up of a large family. *Haematologica* 2016;101:846–52.
39. Moshous D, Martin E, Carpentier W, Lim A, Callebaut I, Canioni D, et al. Whole-exome sequencing identified Coronin-1A deficiency in 3 siblings with immunodeficiency and EBV-associated B-cell lymphoproliferation. *J Allergy Clin Immunol* 2013;131:1594–603.
40. Bandapali OR, Paramasivam N, Giangioffe S, Kumar A, Benisch W, Engert A, et al. Whole genome sequencing reveals DICER1 as a candidate predisposing gene in familial Hodgkin lymphoma. *Int J Cancer* 2018;143:2076–8.
41. Moriyama T, Metzger ML, Wu G, Nishii R, Qian M, Devidas M, et al. Germline genetic variation in ETV6 and risk of childhood acute lymphoblastic leukaemia: a systematic genetic study. *Lancet Oncol* 2015;16:1659–66.
42. Daschkey S, Bienemann K, Schuster V, Kreth HW, Linka RM, Honscheid A, et al. Fatal lymphoproliferative disease in two siblings lacking functional FAAP24. *J Clin Immunol* 2016;36:684–92.
43. Gayden T, Sepulveda FE, Khuong-Quang DA, Pratt J, Valera ET, Garrigue A, et al. Germline HAVCR2 mutations altering TIM-3 characterize subcutaneous panniculitis-like T cell lymphomas with hemaphagocytic lymphohistiocytic syndrome. *Nat Genet* 2018;50:1650–7.
44. Goldin LR, McMaster ML, Rotunno M, Herman SE, Jones K, Zhu B, et al. Whole exome sequencing in families with CLL detects a variant in Integrin  $\beta 2$  associated with disease susceptibility. *Blood* 2016;128:2261–3.
45. Rotunno M, McMaster ML, Boland J, Bass S, Zhang X, Burdett L, et al. Whole exome sequencing in families at high risk for Hodgkin lymphoma: identification of a predisposing mutation in the KDR gene. *Haematologica* 2016;101:853–60.
46. Saarinen S, Kaasinen E, Karjalainen-Lindsberg ML, Vesanen K, Aavikko M, Katainen R, et al. Primary mediastinal large B-cell lymphoma segregating in a family: exome sequencing identified MLL as a candidate predisposition gene. *Blood* 2013;121:3428–30.
47. Shah S, Schrader KA, Waanders E, Timms AE, Vijai J, Meithing C, et al. A recurrent germline PAX5 mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. *Nat Genet* 2013;45:1226–31.
48. McMaster ML, Sun C, Landi MT, Savage SA, Rotunno M, Yang XR, et al. Germline mutations in protection of telomeres 1 in two families with Hodgkin lymphoma. *Br J Haematol* 2018;181:372–7.
49. Hussin J, Sinnett D, Casals F, Idaghmour Y, Bruat V, Saillour V, et al. Rare allelic forms of PRDM9 associated with childhood leukemogenesis. *Genome Res* 2013;23:419–30.
50. Powell BC, Jiang L, Mazny DM, Treviño LR, Dreyer ZE, Strong LC, et al. Identification of TP53 as an acute lymphocytic leukemia susceptibility gene through exome sequencing. *Pediatr Blood Cancer* 2013;60:E1–3.
51. Waanders E, Scheijen B, Jongmans MC, Venselaar H, van Reijmersdal SV, van Dijk AH, et al. Germline activating TYK2 mutations in pediatric patients with two primary acute lymphoblastic leukemia occurrences. *Leukemia* 2017;31:821–8.
52. Roccaro AM, Sacco A, Shi J, Chiarini M, Perilla-Glen A, Manier S, et al. Exome sequencing reveals recurrent germ line variants in patients with familial Waldenström macroglobulinemia. *Blood* 2016;127:2598–606.
53. Lawrie A, Han S, Sud A, Hosking F, Cezard T, Turner D, et al. Combined linkage and association analysis of classical Hodgkin lymphoma. *Oncotarget* 2018;9:20377–85.
54. Speedy HE, Kinnersley B, Chubb D, Broderick P, Law PJ, Litchfield K, et al. Germ line mutations in shelterin complex genes are associated with familial chronic lymphocytic leukemia. *Blood* 2016;128:2319–26.
55. Hirvonen EAM, Pitkänen E, Hemminki K, Aaltonen LA, Kilpivaara O. Whole-exome sequencing identifies novel candidate predisposition genes for familial polycythemia vera. *Hum Genomics* 2017;11:6.
56. Spinella JF, Healy J, Saillour V, Richer C, Cassart P, Ouimet M, et al. Whole-exome sequencing of a rare case of familial childhood acute lymphoblastic leukemia reveals putative predisposing mutations in Fanconi anemia genes. *BMC Cancer* 2015;15:539.
57. Donner I, Katainen R, Kaasinen E, Aavikko M, Sipilä LJ, Pukkala E, et al. Candidate susceptibility variants in angioimmunoblastic T-cell lymphoma. *Fam Cancer* 2019;18:113–9.
58. Valentine MC, Linabery AM, Chasnoff S, Hughes AE, Mallaney C, Sanchez N, et al. Excess congenital non-synonymous variation in leukemia-associated genes in MLL-infant leukemia: a Children’s Oncology Group report. *Leukemia* 2014;28:1235–41.
59. Al-Dewik N, Ben-Omran T, Zayed H, Trujillano D, Kishore S, Rolfs A, et al. Clinical exome sequencing unravels new disease-causing mutations in the myeloproliferative neoplasms: a pilot study in patients from the state of Qatar. *Gene* 2019;689:34–42.
60. Andreeva TV, Tyazhelova TV, Rykalina VN, Gusev FE, Goltsov AY, Zolotareva OI, et al. Whole exome sequencing links dental tumor to an autosomal-dominant mutation in ANO5 gene associated with gnathodiaphyseal dysplasia and muscle dystrophies. *Sci Rep* 2016;6:26440.
61. Heddar A, Fermey P, Coutant S, Angot E, Sabourin JC, Michelin P, et al. Familial solitary chondrosarcoma resulting from germline EXT2 mutation. *Genes Chromosomes Cancer* 2017;56:1280134.
62. Pillay N, Plagnol V, Tarpey PS, Lobo SB, Presneau N, Szuhai K, et al. A common single-nucleotide variant in T is strongly associated with chordoma. *Nat Genet* 2012;44:1185–7.
63. Ronellenfitsch MW, Oh JE, Satomi K, Sumi K, Harter PN, Steinbach JP, et al. CASP9 germline mutation in a family with multiple brain tumors. *Brain Pathol* 2018;28:94–102.
64. De Mariano M, Gallesio R, Chierici M, Furlanello C, Conte M, Garaventa A, et al. Identification of GALNT14 as a novel neuroblastoma predisposition gene. *Oncotarget* 2015;6:26335–46.
65. Wang L, Yamaguchi S, Burstein MD, Terashima K, Chang K, Ng HK, et al. Novel somatic and germline mutations in intracranial germ cell tumours. *Nature* 2014;511:241–5.

66. Andrianova MA, Chetan GK, Sibir MK, McKee T, Merkler D, Narasinga RK, et al. Germline PMS2 and somatic POLE exonuclease mutations cause hypermutability of the leading DNA strand in biallelic mismatch repair deficiency syndrome brain tumours. *J Pathol* 2017;243:331–41.
67. Bainbridge MN, Armstrong GN, Gramatges MM, Bertuch AA, Jhangiani SN, Doddapaneni H, et al. Germline mutations in shelterin complex genes are associated with familial glioma. *J Natl Cancer Inst* 2014;107:384.
68. Smith MJ, O'Sullivan J, Bhaskar SS, Hadfield KD, Poke G, Caird J, et al. Loss-of-function mutations in SMARCE1 cause an inherited disorder of multiple spinal meningiomas. *Nat Genet* 2013;45:295–8.
69. Aavikko M, Li SP, Saarinen S, Alhophuro P, Kaasinen E, Morgunova E, et al. Loss of SUFU function in familial multiple meningioma. *Am J Hum Genet* 2012;91:520–6.
70. Nordfors K, Haapasalo J, Afyounian E, Tuominen J, Annala M, Häyrynen S, et al. Whole-exome sequencing identifies germline mutation in TP53 and ATRX in a child with genomically aberrant AT/RT and her mother with anaplastic astrocytoma. *Cold Spring Harb Mol Case Stud* 2018;4:a002246.
71. Kim YH, Ohta T, Oh JE, Le Calvez-Kelm F, McKay J, Voegelé C, et al. TP53, MSH4, and LATS1 germline mutations in a family with clustering of nervous system tumors. *Am J Pathol* 2014;184:2374–81.
72. Backes C, Harz C, Fischer U, Schmitt J, Ludwig N, Petersen BS, et al. New insights into the genetics of glioblastoma multiforme by familial exome sequencing. *Oncotarget* 2015;6:5918–31.
73. Park DJ, Odefrey FA, Hammet F, Giles GG, Baglietto L, Hopper JL, et al. FAN1 variants identified in multiple-case early-onset breast cancer families via exome sequencing: no evidence for association with risk for breast cancer. *Breast Cancer Res Treat* 2011;130:1043–9.
74. Hilbers FS, Meijers CM, Laros JF, van Galen M, Hoogerbrugge N, Vasen HF, et al. Exome sequencing of germline DNA from non-BRCA1/2 familial breast cancer cases selected on basis of aCGH tumor profiling. *PLoS One* 2013;8:e55732.
75. Radmanesh H, Spethmann T, Enben J, Schümann P, Bhuju S, Geffers R, et al. Assessment of an APOBEC3B truncating mutation, c.783del G, in patients with breast cancer. *Breast Cancer Res Treat* 2017;162:31–7.
76. Tavera-Tapia A, Pérez-Cabornero L, Macías JA, Ceballos MI, Roncador G, de la Hoya M, et al. Almost 2% of Spanish breast cancer families are associated to germline pathogenic mutations in the ATM gene. *Breast Cancer Res Treat* 2017;161:597–604.
77. Vijai J, Topka S, Villano D, Ravichandran V, Maxwell KN, Maria A, et al. A recurrent ERCC3 truncating mutation confers moderate risk for breast cancer. *Cancer Discov* 2016;6:1267–75.
78. Kiiski JI, Pelttari LM, Khan S, Frysteinsdottir ES, Reynisdottir I, Hart SN, et al. Exome sequencing identified FANCM as a susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci U S A* 2014;111:15172–7.
79. Sokolenko AP, Bulanova DR, Iyevleva AG, Aleksakhina SN, Preobrazhenskaya EV, Ivanstov AO, et al. High prevalence of GPRC5A germline mutations in BRCA1-mutant breast cancer patients. *Int J Cancer* 2014;134:2352–8.
80. Lynch H, Wen H, Kim YC, Snyder C, Kinarsky Y, Chen PX, et al. Can unknown predisposition in familial breast cancer be family-specific? *Breast J* 2013;19:520–8.
81. Silvestri V, Zelli V, Valentini V, Rizzolo P, Navazio AS, Coppa A, et al. Whole-exome sequencing and targeted gene sequencing provide insights into the role of PALB2 as a male breast cancer susceptibility gene. *Cancer* 2017;123:210–8.
82. Riahi A, Radmanesh H, Schümann P, Bogdanova N, Geffers R, Meddeb R, et al. Exome sequencing and case-control analyses identify RCC1 as a candidate breast cancer susceptibility gene. *Int J Cancer* 2018;142:2512–7.
83. Cybulski C, Carrot-Zhang J, Kluzniak W, Rivera B, Kashyap A, Wokolorczyk D, et al. Germline RECQL mutations are associated with breast cancer susceptibility. *Nat Genet* 2015;47:643–6.
84. Sun J, Wang Y, Xia Y, Xu Y, Ouyang T, Li J, et al. Mutations in RECQL gene are associated with predisposition to breast cancer. *PLoS Genet* 2015;11:e1005228.
85. Park DJ, Tao K, Le Calvez-Kelm F, Nguyen-Dumont T, Robinot N, Hammet F, et al. Rare mutations in RINT1 predispose carriers to breast and Lynch syndrome-spectrum cancers. *Cancer Discov* 2014;4:804–15.
86. Nguyen-Dumont T, Teo ZL, Hammet F, Roberge A, Mahmoodi M, Tsimiklis H, et al. Is RNASEL.p.Glu265\* a modifier of early-onset breast cancer risk for carriers of high-risk mutations? *BMC Cancer* 2018;18:165.
87. Park DJ, Lesueur F, Nguyen-Dumont T, Pertesi M, Odefrey F, Hammet F, et al. Rare mutations in XRCC2 increase the risk of breast cancer. *Am J Hum Genet* 2012;90:734–9.
88. Ataei-Kachouei M, Nadaf J, Akbari MT, Atri M, Majewski J, Riazalhosseini Y, et al. Double heterozygosity of BRCA2 and STK11 in familial breast cancer detected by exome sequencing. *Iran J Public Health* 2015;44:1348–52.
89. Thompson ER, Doyle MA, Ryland GL, Rowley SM, Choong DY, Tothill RW, et al. Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. *PLoS Genet* 2012;8:31002894.
90. Merdad A, Gari MA, Hussein S, Al-Khayat S, Tashkandi H, Al-Maghrabi J, et al. Characterization of familial breast cancer in Saudi Arabia. *BMC Genomics* 2015;16(Suppl 1):S3.
91. Guo X, Shi J, Cai Q, Shu XO, He J, Wen W, et al. Use of deep whole-genome sequencing data to identify structure risk variants in breast cancer susceptibility genes. *Hum Mol Genet* 2018;27:853–9.
92. Noh JM, Kim J, Cho DY, Choi DH, Park W, Huh SJ. Exome sequencing in a breast cancer family without BRCA mutation. *Radiat Oncol J* 2015;33:149–54.
93. Cybulski C, Lubiński J, Wokolorczyk D, Kluzniak W, Kashyap A, Sopik V, et al. Mutations predisposing to breast cancer in 12 candidate genes in breast cancer patients from Poland. *Clin Genet* 2015;88:366–70.
94. Garcia-Aznarez FJ, Fernandez V, Pita G, Peterlongo P, Dominguez O, de la Hoya M, et al. Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. *PLoS One* 2013;8:e55681.
95. Hamdi Y, Boujemaa M, Ben Rekaya M, Ben Hamda C, Mighri N, El Benna H, et al. Family specific genetic predisposition to breast cancer: results from Tunisian whole exome sequenced breast cancer cases. *J Transl Med* 2018;16:158.
96. Torrezan GT, de Almeida FGDSR, Figueiredo MCP, Barros BDF, de Paula CAA, Valieris R, et al. Complex landscape of germline variants in Brazilian patients with hereditary and early onset breast cancer. *Front Genet* 2018;9:161.
97. Jalkh N, Chouery E, Haidar Z, Khater C, Atallah D, Ali H, et al. Next-generation sequencing in familial breast cancer patients from Lebanon. *BMC Med Genomics* 2017;10:8.
98. Kim YC, Soliman AS, Cui J, Ramadan M, Hablas A, Abouelhoda M, et al. Unique features of germline variation in five Egyptian familial breast cancer families revealed by exome sequencing. *PLoS One* 2017;12:E0167581.
99. Wen H, Kim YC, Snyder C, Xiao F, Fleissner EA, Becirovic D, et al. Family-specific, novel, deleterious germline variants provide a rich resource to identify genetic predispositions for BRCAx familial breast cancer. *BMC Cancer* 2014;14:470.
100. Snape K, Rurak E, Tarpey P, Renwick A, Turnbull C, Seal S, et al. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. *Breast Cancer Res Treat* 2012;134:429–33.
101. Piccolo SR, Hoffman LM, Conner T, Shrestha G, Cohen AL, Marks JR, et al. Integrative analyses reveal signaling pathways underlying familial breast cancer susceptibility. *Mol Syst Biol* 2016;12:860.
102. Bellido F, Sowada N, Mur P, Lázaro C, Pons T, Valdés-Mas R, et al. Association between germline mutations in BRF1, a subunit of the RNA polymerase III transcription complex, and hereditary colorectal cancer. *Gastroenterology* 2018;154:181–94.
103. Seguí N, Mina LB, Lázaro C, Sanz-Pamplona R, Pons T, Navarro M, et al. Germline mutations in FAN1 cause hereditary colorectal cancer by impairing DNA repair. *Gastroenterology* 2015;149:563–6.
104. Segui N, Navarro M, Pineda M, Köger N, Bellido F, González S, et al. Exome sequencing identified MUTYH mutations in a family with colorectal cancer and an atypical phenotype. *Gut* 2015;64:355–6.
105. Nieminen TT, O'Donohue MF, Wu Y, Lohi H, Scherer SW, Paterson AD, et al. Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. *Gastroenterology* 2014;147:595–8.
106. Martín-Morales L, Feldman M, Vershini Z, Garre P, Caldés T, Levy D. SETD6 dominant negative mutation in familial colorectal cancer type X. *Hum Mol Genet* 2017;26:4481–93.
107. DeRycke MS, Gunawardena SR, Middha S, Asmann YW, Schaid DJ, McDonnell SK, et al. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. *Cancer Epidemiol Biomarkers Prev* 2013;22:1239–51.
108. Palles C, Cazier JG, Howarth KM, Domingo E, Jones AM, Broderick P, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet* 2013;45:136–44.

109. Arora S, Yan H, Cho I, Fan HY, Luo B, Gai X, et al. Genetic variants that predispose to DNA double-strand breaks in lymphocytes from a subset of patients with familial colorectal carcinomas. *Gastroenterology* 2015;149:1872–83.
110. de Voer RM, Hahn MM, Weren RD, Mensenkamp AR, Gillissen C, van Zelst-Stams WA, et al. Identification of novel candidate genes for early-onset colorectal cancer susceptibility. *PLoS Genet* 2016;12:31005880.
111. Franch-Expósito S, Esteban-Jurado C, Garre P, Quintanilla I, Duran-Sancho S, Díaz-Gay M, et al. Rare germline copy number variants in colorectal cancer predisposition characterized by exome sequencing analysis. *J Genet Genomics* 2018;45:41–5.
112. Esteban-Jurado C, Franch-Expósito S, Muñoz J, Ocaña T, Carballal S, López-Cerón M, et al. The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer. *Eur J Hum Genet* 2016;24:1501–5.
113. Smith CG, Naven M, Harris R, Colley J, West H, Li N, et al. Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. *Hum Mutat* 2013;34:1026–34.
114. Chubb D, Broderick P, Dobbins SE, Frampton M, Kinnersley B, Penegar S, et al. Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. *Nat Commun* 2016;7:11883.
115. Esteban-Jurado C, Vila-Casadesús M, Garre P, Lozano JJ, Pristoupilova A, Beltran S, et al. Whole-exome sequencing identified rare pathogenic variants in new predisposition genes for familial colorectal cancer. *Genet Med* 2015;17:131–42.
116. Tanskanen T, Gylfe AE, Katainen R, Taipale M, Renkonen-Sinisalo L, Järvinen H, et al. Systematic search for rare variants in Finnish early-onset colorectal cancer patients. *Cancer Genet* 2015;208:35–40.
117. Gylfe AE, Katainen R, Kondelin J, Tanskanen T, Cajuso T, Hänninen U, et al. Eleven candidate susceptibility genes for common familial colorectal cancer. *PLoS Genet* 2013;9:e1003876.
118. Yu L, Yin B, Qu K, Li J, Jin Q, Liu L, et al. Screening for susceptibility genes in hereditary non-polyposis colorectal cancer. *Oncol Lett* 2018;15:9413–9.
119. Zhang JX, Fu L, de Voer RM, Hahn MM, Jin P, Lv CX, et al. Candidate colorectal cancer predisposing gene variants in Chinese early-onset and familial cases. *World J Gastroenterol* 2015;21:4136–49.
120. Khalilipour N, Baranova A, Jebelli A, Heravi-Moussavi A, Bruskin S, Abbaszadegan MR. Familial esophageal squamous cell carcinoma with damaging rare/germline mutations in KCNJ12/KCNJ18 and GPRIN2 genes. *Cancer Genet* 2018;221:46–52.
121. Sherman SK, Maxwell JE, Qian Q, Bellizzi AM, Braun TA, Iannettoni MD, et al. Esophageal cancer in a family with hamartomatous tumors and germline PTEN frameshift and SMAD7 missense mutations. *Cancer Genet* 2015;208:41–6.
122. Donner I, Katainen R, Tanskanen T, Kaasinen E, Avikko M, Ovaska K, et al. Candidate susceptibility variants for esophageal squamous cell carcinoma. *Genes Chromosomes Cancer* 2017;56:453–9.
123. Forouzanfar N, Baranova A, Milanizadeh S, Heravi-Moussavi A, Jebelli A, Abbaszadegan MR. Novel candidate genes may be possible predisposing factors revealed by whole exome sequencing in familial esophageal squamous cell carcinoma. *Tumour Biol* 2017;39:1010428317699115.
124. Vogelaar IP, van der Post RS, van Krieken JHJ, Spruijt L, van Zelst-Stams WA, Kets CM, et al. Unraveling genetic predisposition to familial or early onset gastric cancer using germline whole-exome sequencing. *Eur J Hum Genet* 2017;25:1246–52.
125. Calvete O, Reyes J, Zuñiga S, Paumard-Hernández B, Fernández V, Bujanda L, et al. Exome sequencing identifies ATP4A gene as responsible of an atypical familial type I gastric neuroendocrine tumour. *Hum Mol Genet* 2015;24:2914–22.
126. Majewski JJ, Kluij I, Cats A, Scerri TS, de Jong D, Kluij R, et al. An  $\alpha$ -E-catenin (CTNNA1) mutation in hereditary diffuse gastric cancer. *J Pathol* 2013;229:621–9.
127. Donner I, Kiviluoto T, Ristimäki A, Aaltonen LA, Vahteristo P. Exome sequencing reveals three novel candidate predisposition genes for diffuse gastric cancer. *Fam Cancer* 2015;14:241–6.
128. Sahasrabudhe R, Lott P, Bohorquez M, Toal T, Estrada AP, Suarez JJ, et al. Germline mutations in PALB2, BRCA1, and RAD1C, which regulate DNA recombination repair, in patients with gastric cancer. *Gastroenterology* 2017;152:983–6.
129. Fewings E, Larionov A, Redman J, Goldgraben MA, Scarth J, Richardson S, et al. Germline pathogenic variants in PALB2 and other cancer-predisposing genes in families with hereditary diffuse gastric cancer without CHD1 mutation: a whole-exome sequencing study. *Lancet Gastroenterol Hepatol* 2018;3:489–98.
130. Thutkawkorapin J, Picelli S, Kontham V, Liu T, Nilsson D, Lindblom A. Exome sequencing in one family with gastric and rectal cancer. *BMC Genet* 2017;17:41.
131. Comino-Méndez I, García-Aznárez FJ, Schiavi F, Landa I, Leandro-García LJ, Letón R, et al. Exome sequencing identified MAX mutations as a cause of hereditary pheochromocytoma. *Nat Genet* 2011;43:663–7.
132. Cao M, Sun F, Huang X, Dai J, Cui B, Ning G. Analysis of the inheritance pattern of a Chinese family with pheochromocytomas through whole exome sequencing. *Gene* 2013;526:164–9.
133. Channir HI, van Overeem Hansen T, Andreassen S, Yde CW, Kiss K, Charabi BW. Genetic characterization of adenoid cystic carcinoma of the minor salivary glands: a potential familial occurrence in first-degree relatives. *Head Neck Pathol* 2017;11:546–51.
134. Sasaki MM, Slol AD, Bao R, Rhodes LV, Chambers R, Vokes EE, et al. Integrated genomic analysis suggests MLL3 is a novel candidate susceptibility gene for familial nasopharyngeal carcinoma. *Cancer Epidemiol Biomarkers Prev* 2015;24:1222–8.
135. Dai W, Zheng H, Cheung AK, Tang CS, Ko JM, Wong BW, et al. Whole-exome sequencing identified MST1R as a genetic susceptibility gene in nasopharyngeal carcinoma. *Proc Natl Acad Sci U S A* 2016;113:3317–22.
136. Das R, Kundu S, Laskar S, Choudhury Y, Ghosh SK. Assessment of DNA repair susceptibility genes identified by whole exome sequencing in head and neck cancer. *DNA Repair* 2018;66–67:50–63.
137. Sei Y, Zhao X, Forbes J, Szymczak S, Li Q, Trivedi A, et al. A hereditary form of small intestinal carcinoid associated with a germline mutation in inositol polyphosphate multikinase. *Gastroenterology* 2015;149:67–78.
138. Dumanski JP, Rasi C, Björklund P, Davies H, Ali AS, Grönberg M, et al. A MUTYH germline mutation is associated with small intestinal neuroendocrine tumors. *Endocr Relat Cancer* 2017;24:427–45.
139. Popova T, Hebert L, Jacquemin V, Gad S, Caux-Moncoutier V, Dubois-d'Enghien C, et al. Germline BAP1 mutations predispose to renal cell carcinomas. *Am J Hum Genet* 2013;92:974–80.
140. Hanks S, Perdeaux ER, Seal S, Ruark E, Mahamdallie SS, Murray A, et al. Germline mutations in the PAF1 complex gene CTR9 predispose to Wilms tumour. *Nat Commun* 2014;5:4396.
141. Mahamdallie SS, Hanks S, Karlin KL, Zachariou A, Perdeaux ER, Ruark E, et al. Mutations in the transcriptional repressor REST predispose to Wilms tumor. *Nat Genet* 2015;47:1471–4.
142. Palculict TB, Ruteshouser EC, Fan Y, Wang W, Strong L, Huff V. Identification of germline DICER1 mutations and loss of heterozygosity in familial Wilms tumour. *J Med Genet* 2016;53:385–8.
143. Caruso S, Calderaro J, Letouzé E, Nault JC, Couchy G, Goulay A, et al. Germline and somatic DICER1 mutations in familial and sporadic liver tumors. *J Hepatol* 2017;66:734–42.
144. Kukita Y, Okami J, Yoneda-Kato N, Nakamae I, Kawabata T, Higashiyama M, et al. Homozygous inactivation of CHEK2 is linked to a familial case of multiple primary lung cancer with accompanying cancers in other organs. *Cold Spring Harb Mol Case Study* 2016;2:a001032.
145. Yamamoto H, Higasa K, Sakaguchi M, Shien K, Soh J, Ichimura K, et al. Novel germline mutation in the transmembrane domain of HER2 in familial lung adenocarcinomas. *J Natl Cancer Inst* 2014;106:djt338.
146. Lusk CM, Wenzlaff AS, Dyson G, Purrington KS, Watzka D, Land S. Whole-exome sequencing reveals genetic variability among lung cancer cases subphenotyped for emphysema. *Carcinogenesis* 2016;37:139–44.
147. Tode N, Kikuchi T, Sakakibara T, Hirano T, Inoue A, Ohkouchi S, et al. Exome sequencing deciphers a germline MET mutation in familial epidermal growth factor receptor-mutant lung cancer. *Cancer Sci* 2017;108:1263–70.
148. Xiong D, Wang Y, Kupert E, Simpson C, Pinney SM, Gaba CR, et al. A recurrent mutation in PARK2 is associated with familial lung cancer. *Am J Hum Genet* 2015;96:301–8.
149. Chen HY, Yu SL, Ho BC, Su KY, Hsu YC, Chang CS, et al. R331W missense mutation of oncogene YAP1 is a germline risk allele for lung adenocarcinoma with medical actionability. *J Clin Oncol* 2015;33:2303–10.
150. He Y, Li Y, Qui Z, Zhou B, Shi S, Zhang K, et al. Identification and validation of PROM1 and CRTC2 mutations in lung cancer patients. *Mol Cancer* 2014;13:19.

151. Tomoshige K, Matsumoto K, Tsuchiya T, Oikawa M, Miyazaki T, Yamasaki N, et al. Germline mutations causing familial lung cancer. *J Hum Genet* 2015;69:597–603.
152. Kanwal M, Ding XJ, Ma ZH, Li LW, Wang P, Chen Y, et al. Characterization of germline mutations in familial lung cancer from the Chinese population. *Gene* 2018;641:94–104.
153. Luo W, Tian P, Wang Y, Xu H, Chen L, Tang C, et al. Characteristics of genomic alterations of lung adenocarcinoma in young never-smokers. *Int J Cancer* 2018;143:1696–705.
154. Esai Selvan M, Klein RJ, Gümüş ZH. Rare, pathogenic germline variants in Fanconi Anemia genes increase risk for squamous lung cancer. *Clin Cancer Res* 2019;25:1517–25.
155. Renieri A, Mencarelli MA, Cetta F, Baldassarri M, Mari F, Furini S, et al. Oligogenic germline mutations identified in early non-smokers lung adenocarcinoma patients. *Lung Cancer* 2014;85:168–74.
156. Donner I, Katainen R, Sipilä LJ, Aavikko M, Pukkala E, Aaltonen LA. Germline mutations in young non-smoking women with lung adenocarcinoma. *Lung Cancer* 2018;122:76–82.
157. Dębniak T, Scott RJ, Lea RA, Górski B, Masojć B, Cybulski C, et al. Founder mutations for early onset melanoma as revealed by whole exome sequencing suggests that this is not associated with the increasing incidence of melanoma in Poland. *Cancer Res Treat* 2019;51:337–44.
158. Aoude LG, Wadt K, Bojesen A, Crüger D, Borg A, Trent JM, et al. A BAP1 mutation in a Danish family predisposes to uveal melanoma and other cancers. *PLoS One* 2013;8:e72144.
159. Teerlink CC, Huff C, Stevens J, Yu Y, Holmen SL, Silvis MR, et al. A nonsynonymous variant in the GOLM1 gene in cutaneous malignant melanoma. *J Natl Cancer Inst* 2018;110:1380–5.
160. Aoude LG, Heitzer E, Johansson P, Gartside M, Wadt K, Pritchard AL, et al. POLE mutations in families predisposed to cutaneous melanoma. *Fam Cancer* 2015;14:621–8.
161. Shi J, Yang XR, Ballew B, Rotunno M, Calista D, Fargnoli MC, et al. Rare missense variants in POT1 predisposes to familial cutaneous malignant melanoma. *Nat Genet* 2014;46:482–6.
162. Robles-Espinoza CD, Harland M, Ramsay AJ, Aoude LG, Quesada V, Ding Z, et al. POT1 loss-of-function variants predispose to familial melanoma. *Nat Genet* 2014;46:478–81.
163. Wadt KA, Aoude LG, Golmard L, Hansen TV, Sastre-Garau X, Hayward NK, et al. Germline RAD51B truncating mutation in a family with cutaneous melanoma. *Fam Cancer* 2015;14:337–40.
164. Benfodda M, Gazal S, Descamps V, Basset-Seguín N, Deschamps L, Thomas L, et al. Truncating mutations of TP53AIP1 gene predispose to cutaneous melanoma. *Genes Chromosomes Cancer* 2018;57:294–303.
165. Ibarrola-Villava M, Kumar R, Nagore E, Benfodda M, Guedj M, Gazal S, et al. Genes involved in the WNT and vesicular trafficking pathways are associated with melanoma predisposition. *Int J Cancer* 2015;136:2109–19.
166. Artomov M, Stratigos AJ, Kim I, Kumar R, Lauss M, Reddy BY, et al. Rare variant, gene-based associations study of hereditary melanoma using whole-exome sequencing. *J Natl Cancer Inst* 2017;109:djx083.
167. Yu Y, Hu H, Chen JS, Hu F, Fowler J, Scheet P, et al. Integrated case-control and somatic-germline interaction analyses of melanoma susceptibility genes. *Biochim Biophys Acta Mol Basis Dis* 2018;1864:2247–54.
168. Franceschi S, Spugnese L, Aretini P, Lessi F, Scarpitta R, Galli A, et al. Whole-exome analysis of a Li-Fraumeni family trio with a novel TP53 PRD mutation and anticipation profile. *Carcinogenesis* 2017;38:938–43.
169. Wadt K, Choi J, Chung JY, Kiilgaard J, Heedgaard S, Drzewiecki KT, et al. A cryptic BAP1 splice mutation in a family with uveal and cutaneous melanoma, and paraganglioma. *Pigment Cell Melanoma Res* 2012;25:815–8.
170. Wilson TL, Hattangady N, Lerario AM, Williams C, Koeppe E, Quinonez S, et al. A new POT1 germline mutation-expanding the spectrum of POT1-associated cancers. *Fam Cancer* 2017;16:561–6.
171. Li WD, Li QR, Xu SN, Wei FJ, Ye ZJ, Cheng JK, et al. Exome sequencing identifies an MLL3 gene germline mutation in a pedigree of colorectal cancer and acute myeloid leukemia. *Blood* 2013;121:1478–9.
172. Rohlin A, Zagoras T, Nilsson S, Lundstam U, Wahlström J, Hultén L, et al. A mutation in POLE predisposing to a multi-tumour phenotype. *Int J Oncol* 2014;45:77–81.
173. Hansen MF, Johansen J, Bjornevold I, Sylvander AE, Steinsbekk KS, Sætrom P, et al. A novel POLE mutation associated with cancers of colon, pancreas, ovaries and small intestine. *Fam Cancer* 2015;14:437–48.
174. Wei C, Peng B, Han Y, Chen WV, Rother J, Tomlinson GE, et al. Mutations of HNRNPA0 and WIF1 predispose members of a large family to multiple cancers. *Fam Cancer* 2015;14:297–306.
175. Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MD, et al. Patterns and functional implications of rare germline variants across 12 cancer types. *Nat Commun* 2015;6:10086.
176. Huang KL, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, et al. Pathogenic germline variants in 10,389 adult cancers. *Cell* 2018;173:355–70.
177. Diets IJ, Waanders E, Ligtenberg MJ, van Bladel DAG, Kamping EJ, Hoogerbrugge PM, et al. High yield of pathogenic germline mutations causative or likely causative of the cancer phenotype in selected children with cancer. *Clin Cancer Res* 2018;24:1594–603.
178. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, et al. Germline mutations in predisposition genes in pediatric cancer. *N Engl J Med* 2015;373:2336–46.
179. Dicks E, Song H, Ramus SJ, Oudenhove EV, Tyrer JP, Intermaggio MP, et al. Germline whole exome sequencing and large-scale replication identified FANCM as a likely high grade serous ovarian cancer susceptibility gene. *Oncotarget* 2017;8:50930–40.
180. Witkowski L, Carrot-Zhang J, Albrecht S, Fahiminiya S, Hamel N, Tomiak E, et al. Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. *Nat Genet* 2014;46:438–43.
181. Stafford JL, Dyson G, Levin NK, Chaudhry S, Rosati R, Kalpage H, et al. Reanalysis of BRCA1/2 negative high-risk ovarian cancer patients reveals novel germline risk loci and insights into missing heritability. *PLoS One* 2017;12:e0178450.
182. Kanchi KL, Johnson KJ, Lu C, McLellan MD, Leiserson MD, Wendl MC, et al. Integrated analysis of germline and somatic variants in ovarian cancer. *Nat Commun* 2014;5:3156.
183. Määttä K, Rantaperö T, Lindström A, Nykter M, Kankuri-Tammilehto M, Laasanen SL, et al. Whole-exome sequencing of Finnish hereditary breast cancer families. *Eur J Hum Genet* 2016;25:85–93.
184. Lu HM, Li S, Black MH, Lee S, Hoiness R, Wu S, et al. Association of breast and ovarian cancers with predisposition genes identified by large-scale sequencing. *JAMA Oncol* 2019;5:51–7.
185. Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, et al. ATM mutations in patients with hereditary pancreatic cancer. *Cancer Discov* 2012;2:41–6.
186. Grant RC, Denroche RE, Borgida A, Virtanen C, Cook N, Smith AL, et al. Exome-wide association study of pancreatic cancer risk. *Gastroenterology* 2018;154:719–22.
187. Grant RC, Al-Sukhni W, Borgida AE, Holter S, Kanji ZS, McPherson T, et al. Exome sequencing identifies nonsegregating nonsense ATM and PALB2 variants in familial pancreatic cancer. *Hum Genomics* 2013;7:11.
188. Smith AL, Alirezaie N, Connor A, Chan-Seng-Yue M, Grant R, Selander I, et al. Candidate DNA repair susceptibility genes identified by exome sequencing in high-risk pancreatic cancer. *Cancer Lett* 2016;370:302–12.
189. Roberts NJ, Norris AL, Petersen GM, Bondy ML, Brand R, Gallinger S, et al. Whole genome sequencing defines the genetic heterogeneity of familial pancreatic cancer. *Cancer Discov* 2016;6:166–75.
190. Yang XR, Rotunno M, Xiao Y, Ingvar C, Helgadottir H, Pastorino L, et al. Multiple rare variants in high-risk pancreatic cancer-related genes may increase risk for pancreatic cancer in a subset of patients with and without germline CDKN2A mutations. *Hum Genet* 2016;135:1241–12.
191. Rand KA, Rohland N, Tandon A, Stram A, Sheng X, Do R, et al. Whole-exome sequencing of over 4100 men of African ancestry and prostate cancer risk. *Hum Mol Genet* 2016;25:371–81.
192. Fitzgerald LM, Kumar A, Boyle EA, Zhang Y, McIntosh LM, Kolb S, et al. Germline missense variants in the BTNL2 gene are associated with prostate cancer susceptibility. *Cancer Epidemiol Biomarkers Prev* 2013;22:1520–8.
193. Mijuskovic M, Saunders EJ, Leongamornlert DA, Wakerell S, Whitmore I, Dadaev T, et al. Rare germline variants in DNA repair genes and the angiogenesis pathway predispose prostate cancer patients to develop metastatic disease. *Br J Cancer* 2018;119:96–104.
194. Hayano T, Matsui H, Nakaoka H, Ohtake N, Hosomichi K, Suzuki K, et al. Germline variants of prostate cancer in Japanese families. *PLoS One* 2016;11:e0164233.

195. Karyadi DM, Geybels MS, Karlins E, Decker B, McInosh L, Hutchinson A, et al. Whole exome sequencing in 75 high-risk families with validation and replication in independent case-control studies identifies TANGO2, OR5H14, and CHAD as new prostate cancer susceptibility genes. *Oncotarget* 2017;8:1495–507.
196. Johnson AM, Zuhlke KA, Plotts C, McDonnell SK, Middha S, Riska SM, et al. Mutational landscape of candidate genes in familial prostate cancer. *Prostate* 2014;74:1371–8.
197. Hart SN, Ellingson MS, Schahl K, Vedell PT, Carlson RE, Sinnwell JP, et al. Determining the frequency of pathogenic germline variants from exome sequencing in patients with castrate-resistant prostate cancer. *BMJ Open* 2016;6:e010332.
198. Hunter SM, Rowley SM, Clouston DK, Con Fab Investigators Li J, Lupat R, et al. Searching for candidate genes in familial BRCA mutation carriers with prostate cancer. *Urol Oncol* 2016;34:120.
199. Koboldt DC, Kanchi KL, Gui B, Larson DE, Fulton RS, Isaacs WB, et al. Rare variation in TET2 is associated with clinically relevant prostate carcinoma in African Americans. *Cancer Epidemiol Biomarkers Prev* 2016;25:1456–63.
200. Nicolas E, Arora S, Zhou Y, Serebriiskii IG, Andrade MD, Handorf ED, et al. Systematic evaluation of underlying defects in DNA repair as an approach to case-only assessment of familial prostate cancer. *Oncotarget* 2015;6:39614–33.
201. Calvete O, Martinez P, Garcia-Pavia P, Benítez-Buelga C, Paumard-Hernández B, Fernandez V, et al. A mutation in the POT1 gene is responsible for cardiac angiosarcoma in TP53-negative Li-Fraumeni-like families. *Nat Commun* 2015;6:8383.
202. Witkowski L, Lalonde E, Zhang J, Albrecht S, Hamel N, Cavallone L, et al. Familial rhabdoid tumor ‘avant la lettre’—from pathology review to exome sequencing and back again. *J Pathol* 2013;231:35–43.
203. Aavikko M, Kaasinen E, Nieminen JK, Byun M, Donner I, Mancuso R, et al. Whole-genome sequencing identifies STAT4 as a putative susceptibility gene in classic Kaposi sarcoma. *J Infect Dis* 2015;211:1842–51.
204. Jouenne F, Chauvot de Beauchene I, Bollaert E, Avril MF, Caron O, Ingster O, et al. Germline CDKN2A/P16INK4A mutations contribute to genetic determinism of sarcoma. *J Med Genet* 2017;54:607–12.
205. Brohl AS, Patidar R, Turner CE, Wen X, Song YK, Wei JS, et al. Frequent inactivating germline mutations in DNA repair genes in patients with Ewing sarcoma. *Genet Med* 2017;19:955–8.
206. Litchfield K, Loveday C, Levy M, Dudakia D, Rapley E, Nsengimana J, et al. Large-scale sequencing of testicular germ cell tumour (TGCT) cases excludes major TGCT predisposition gene. *Eur Urol* 2018;73:828–31.
207. Truscott L, Gell J, Chang VY, Lee H, Strom SP, Pillai R, et al. Novel association of familial testicular germ cell tumor and autosomal dominant polycystic kidney disease with PKD1 mutation. *Pediatr Blood Cancer* 2017;64:100–2.
208. Paumard-Hernández B, Calvete O, Inglada Pérez L, Tejero H, Al-Hahrour F, Pita G, et al. Whole exome sequencing identified PLEC, EXO5 and DNAH7 as novel susceptibility genes in testicular cancer. *Int J Cancer* 2018;143:1954–62.
209. Litchfield K, Levy M, Dudakia D, Proszek P, Shipley C, Basten S, et al. Rare disruptive mutations in ciliary function genes contribute to testicular cancer susceptibility. *Nat Commun* 2016;7:13840.
210. Liu C, Yu Y, Yin G, Zhang J, Wen W, Ruan X, et al. C14orf93 (RTFC) is identified as a novel susceptibility gene for familial nonmedullary thyroid cancer. *Biochem Biophys Res Commun* 2017;482:590–6.
211. Gara SK, Jia L, Merino MJ, Agarwal SK, Zhang L, Cam M, et al. Germline HAP2 mutation causing familial nonmedullary thyroid cancer. *N Engl J Med* 2015;373:448–55.
212. Ye F, Gao H, Xiao L, Zuo Z, Liu Y, Zhao Q, et al. Whole exome and target sequencing identified MAP2K5 as novel susceptibility gene for familial non-medullary thyroid carcinoma. *Int J Cancer* 2019;144:1321–30.
213. Sponziello M, Benvenuti S, Gentile A, Pecce V, Rosignolo F, Virzi AR, et al. Whole exome sequencing identifies a germline MET mutation in two siblings with hereditary wild-type RET medullary thyroid cancer. *Hum Mutat* 2018;39:371–7.
214. Ikeda Y, Kiyotani K, Yew PY, Kato T, Tamura K, Yap KL, et al. Germline PARP4 mutations in patients with primary thyroid and breast cancers. *Endocr Relat Cancer* 2016;23:171–9.
215. Tomsic J, He H, Akagi K, Liyanarachchi S, Pan Q, Bertani B, et al. A germline mutation in SRRM2, a splicing factor gene, is implicated in papillary thyroid carcinoma predisposition. *Sci Rep* 2015;5:10566.
216. Ni Y, Seballos S, Fletcher B, Romigh T, Yehia L, Mester J, et al. Germline compound heterozygous poly-glutamine deletion in USF3 may be involved in predisposition to heritable and sporadic epithelial thyroid carcinoma. *Hum Mol Genet* 2017;26:243–57.
217. Pinheiro M, Drigo SA, Tonhosolo R, Andrade SCS, Marchi FA, Jurisica I, et al. HAP2 p.G534E variant in patients with family history of thyroid and breast cancer. *Oncotarget* 2017;8:40896–905.
218. Ernst C, Hahnen E, Engel C, Nothnagel M, Weber J, Schmutzler RK, et al. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Med Genomics* 2018;11:35.
219. Leong IU, Stuckey A, Lai D, Skinner JR, Love DR. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. *BMC Med Genet* 2015;16:34.
220. McCarthy DJ, Humburg P, Kanapin A, Rivas MA, Gaulton K, Cazier JB, et al. Choice of transcripts and software has a large effect on variant annotation. *Genome Med* 2014;6:26.
221. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 2013;15:733–47.
222. Landrum MJ, Kattman BL. ClinVar at five years: delivering on the promise. *Hum Mutation* 2018;39:1623–30.
223. Melchor L, Benítez J. The complex genetic landscape of familial breast cancer. *Hum Genet* 2013;132:845–63.
224. Kiezun A, Garimella K, Do R, Stitzel NO, Neale BM, McLaren PJ, et al. Exome sequencing and the genetic basis of complex traits. *Nat Genet* 2012;44:623–30.
225. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;12:745–55.
226. Lappalainen T, Scott AJ, Brandt M, Hallet IM. Genomic analysis in the age of human genome sequencing. *Cell* 2019;177:70.
227. Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, et al. Exome sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. *Science* 2009;324:217.
228. Saarinen S, Aavikko M, Aittomäki K, Launonen V, Lehtonen R, Franssila K, et al. Exome sequencing reveals germline NPAT mutation as a candidate risk factor for Hodgkin lymphoma. *Blood* 2011;118:493–8.