



Compound Heterozygous Variants in Pediatric Cancers: A Systematic Review

Dustin B. Miller and Stephen R. Piccolo*

Department of Biology, Brigham Young University, Provo, UT, United States

OPEN ACCESS

Edited by:

Daria Chudakova,
The University of Auckland,
New Zealand

Reviewed by:

Silvia R. Rogatto,
University of Southern
Denmark, Denmark
Thomas LaFramboise,
Case Western Reserve University,
United States

*Correspondence:

Stephen R. Piccolo
stephen_piccolo@byu.edu

Specialty section:

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Genetics

Received: 14 February 2020

Accepted: 20 April 2020

Published: 19 May 2020

Citation:

Miller DB and Piccolo SR (2020)
Compound Heterozygous Variants in
Pediatric Cancers: A Systematic
Review. *Front. Genet.* 11:493.
doi: 10.3389/fgene.2020.00493

A compound heterozygous (*CH*) variant is a type of germline variant that occurs when each parent donates one alternate allele and these alleles are located at different loci within the same gene. Pathogenic germline variants have been identified for some pediatric cancer types but in most studies, *CH* variants are overlooked. Thus, the prevalence of pathogenic *CH* variants in most pediatric cancer types is unknown. We identified 26 studies (published between 1999 and 2019) that identified a *CH* variant in at least one pediatric cancer patient. These studies encompass 21 cancer types and have collectively identified 25 different genes in which a *CH* variant occurred. However, the sequencing methods used and the number of patients and genes evaluated in each study were highly variable across the studies. In addition, methods for assessing pathogenicity of *CH* variants varied widely and were often not reported. In this review, we discuss technologies and methods for identifying *CH* variants, provide an overview of studies that have identified *CH* variants in pediatric cancer patients, provide insights into future directions in the field, and give a summary of publicly available pediatric cancer sequencing data. Although considerable insights have been gained over the last 20 years, much has yet to be learned about the involvement of *CH* variants in pediatric cancers. In future studies, larger sample sizes, more pediatric cancer types, and better pathogenicity assessment and filtering methods will be needed to move this field forward.

Keywords: pediatric cancer, germline variants, compound heterozygosity, variant pathogenicity assessment, genetic analysis of complex diseases

INTRODUCTION

Each year worldwide, ~300,000 children under the age of 14 are diagnosed with cancer (Sweet-Cordero and Biegel, 2019). Since the 1970s, 5-year survival rates for pediatric cancer patients have steadily increased and are presently over 80% (Phillips et al., 2015). Despite improvements in treatments and survival rates, the causes of most pediatric cancers are still relatively unknown (American Cancer Society¹). Recent large-scale studies have helped elucidate the involvement of germline mutations in pediatric cancer development. For example, a 2015 study analyzed mutations across 565 known, cancer-associated genes for 1,120 pediatric cancer patients and found that 8.5% of the patients had an identifiable, pathogenic germline mutation in at least one of these genes (Zhang et al., 2015). Similarly, a 2016 study identified 10% of pediatric cancer patients as having

¹American Cancer Society Risk Factors and Causes of Childhood Cancer. Available online at: <https://www.cancer.org/cancer/cancer-in-children/risk-factors-and-causes.html> (accessed March 29, 2019).

a mutation in a cancer predisposition gene (Parsons et al., 2016). These studies highlight that heritable predisposition does play a critical role in many pediatric cancer cases (Dean and Farmer, 2017). However, these studies also emphasize the need to elucidate additional types of rare, germline variants associated with pediatric cancers. In this review, we focus on *compound heterozygous (CH)* variants, a type of germline variant that has been understudied in pediatric cancers.

CH variants occur when each parent donates one alternate allele and when the alleles are located at different loci within the same gene (**Figure 1**; Kamphans et al., 2013). *CH* variants are particularly relevant to certain types of genes, such as tumor suppressors, where loss-of-function variations are often recessive (Wang et al., 2018). Tumor suppressor genes are involved in inhibiting cell division, initiating apoptosis, repairing DNA damage, and suppressing metastasis. When tumor suppressor genes lose function, tumors can arise and existing tumors can become more aggressive (Guo et al., 2014; Wang et al., 2018). Commonly, researchers identify cases where two non-reference alleles at a given genomic locus have been inherited, one from each parent (homozygosity). But in other cases, an individual may inherit two defective copies of a tumor suppressor gene—one from each parent, with defects at different loci in the gene—thereby resulting in no functional copies of the tumor suppressor gene and potentially an increased susceptibility to cancer development (Weinberg, 2013).

Sanger sequencing has classically been used to identify *CH* variants. In this approach, DNA is collected from the patient and his/her parents, and visual analysis of an electropherogram helps to confirm Mendelian inheritance of a *CH* variant (Piane et al., 2016; Nafisinia et al., 2017). These technologies are most relevant when the researcher wishes to examine one or a few genes. In recent years, detection of *CH* variants at a larger scale has become feasible with the advent of next-generation sequencing (NGS).

NGS technologies (e.g., Illumina, PacBio) can sequence DNA in a high-throughput manner and thus allow for the examination of many genes in a single run (Tewhey et al., 2011). In order to use NGS data for identifying *CH* variants, sequencing data must be phased. The process of phasing estimates which chromosome (haplotype) the nucleotides are located on, thereby helping to distinguish between maternally and paternally inherited variants (Choi et al., 2018). If it is known before sequencing that haplotype information will be needed, a laboratory-based, haplotype-estimation method can be used, such as Linked-Read technology by 10X Genomics (Zheng et al., 2016). This approach uses micro-droplet-based dilution to compartmentalize DNA in a random manner and uses a high number of distinct barcodes. This method prevents the partitioned DNA molecules from originating from the same genomic loci. Alternatively, if NGS libraries have been prepared without regard to phase, computer-based phasing algorithms can be used to estimate haplotypes (Browning and Browning, 2007; Delaneau et al., 2013; Loh et al., 2016). These algorithms estimate a patient's haplotypes from genotype data after making inferences from a population-based reference panel and/or inferences from parental inheritance patterns.

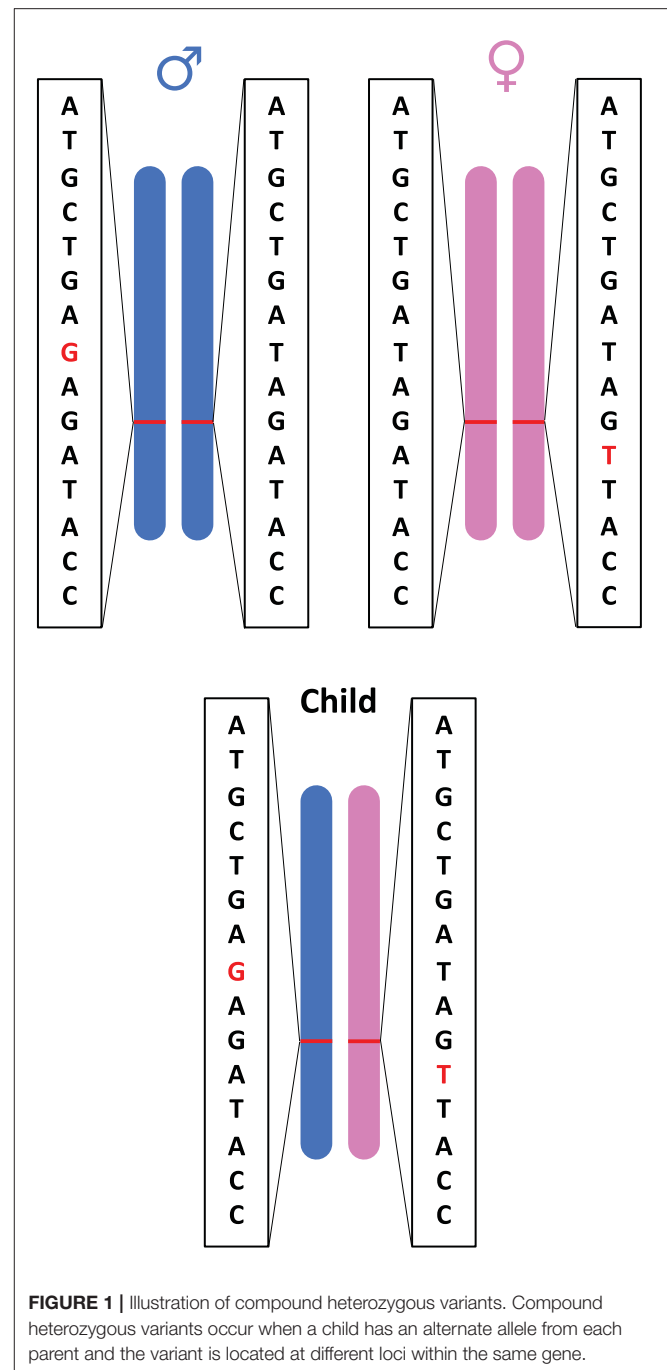
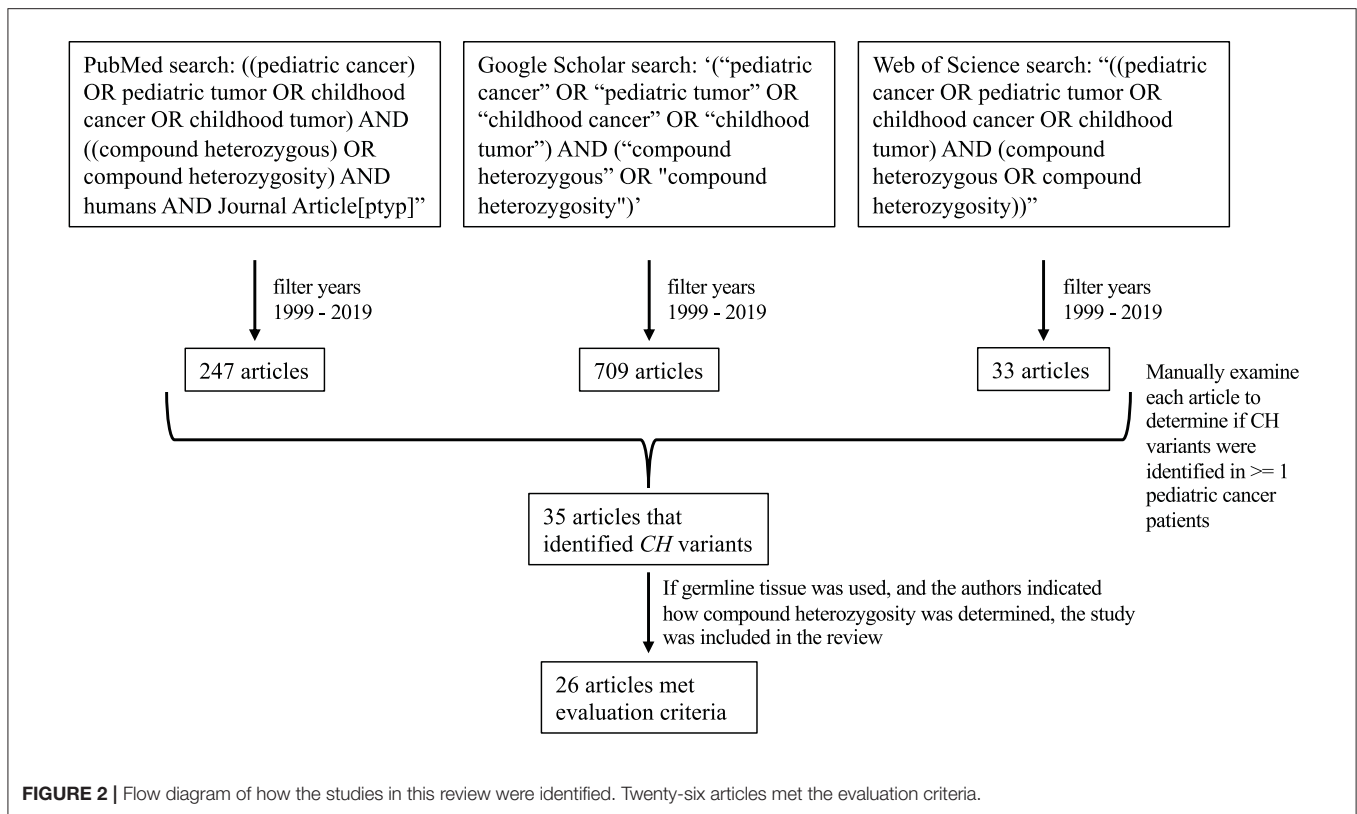


FIGURE 1 | Illustration of compound heterozygous variants. Compound heterozygous variants occur when a child has an alternate allele from each parent and the variant is located at different loci within the same gene.

Assessment of *CH* variant pathogenicity can be accomplished using trusted variant databases, published literature, functional studies, and predictive algorithms (Richards et al., 2015). While each assessment method can be useful, each has unique challenges that may lead to poor pathogenicity assessment. Databases may lack validation data, contain outdated information, or be based on small sample sizes. Published literature may reflect a poor study design or a limited sample size. Functional studies aim to understand the downstream effects of genetic variants (Rodenburg, 2018). For example, gene



rescue assays seek to determine whether introducing the wild-type allele into patient derived cells “rescues” the phenotype. However, functional studies may not reflect the true biological environment or may not take full pathways into consideration (Richards et al., 2015). Common variant prediction algorithms, such as *SIFT* and *PolyPhen-2*, use nucleotide sequence homology (among other parameters) to predict whether protein function is affected by an amino acid substitution (Ng and Henikoff, 2003; Adzhubei et al., 2010). Predictive algorithms often have low specificity for missense variant prediction (Richards et al., 2015). Therefore, in most cases, it is important to use multiple programs for pathogenicity assessment (Niroula and Vihinen, 2019).

In the following sections, we provide an overview of studies that have identified *CH* variants in pediatric-cancer patients, provide insights into future directions in the field, and give a summary of available pediatric cancer sequencing data.

METHODS

Literature Search

On March 20, 2020, we searched PubMed, Google Scholar, and Web of Science for “((pediatric cancer) OR pediatric tumor OR childhood cancer OR childhood tumor) AND ((compound heterozygous) OR compound heterozygosity) AND humans AND Journal Article[ptyp]”, “(“pediatric cancer” OR “pediatric tumor” OR “childhood cancer” OR “childhood tumor”) AND (“compound heterozygous” OR “compound heterozygosity”),

and “((pediatric cancer OR pediatric tumor OR childhood cancer OR childhood tumor) AND (compound heterozygous OR compound heterozygosity)),” respectively. The above searches were also filtered to be inclusive of studies published between 1999 and 2019. Based on these criteria, 247, 709, and 33 results were obtained from PubMed, Google Scholar, and Web of Science, respectively. Of the returned results, we examined each article’s abstract (and full article text when necessary) to determine whether the researchers had identified *CH* variants in one or more pediatric cancer patients; we identified 35 total articles that met these criteria (Figure 2).

Evaluation Criteria

We further evaluated the 35 articles to identify whether the authors had used germline tissue for DNA sequencing and described methods for assessing compound heterozygosity. If germline tissue was used for sequencing, and if the authors indicated how compound heterozygosity was determined, we included the study. Of the 35 studies identified, 26 met these criteria and are described in this review (Figure 2). For each of the 26 studies, we obtained additional information such as tumor type, the number of *CH* variants identified and the genes in which they were identified, the number of patients per cancer type that were included in the study, how compound heterozygosity was determined (i.e., whether through Mendelian inheritance or phasing), the type of sequencing technology used, the use of parental sequence data, and how variant pathogenicity was evaluated (Supplementary Data 1: T1).

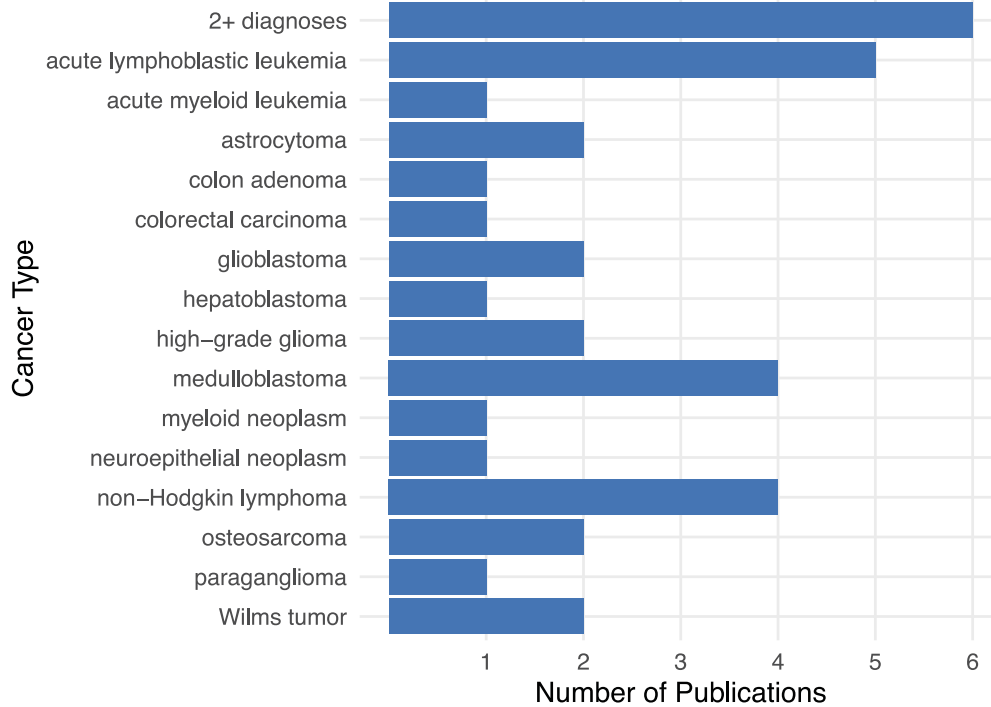


FIGURE 3 | The number of publications per cancer type pertaining to *CH* variants in pediatric cancer. The literature on *CH* variants has covered a wide range of cancer types, especially acute lymphoblastic leukemia, non-Hodgkin's lymphoma, and medulloblastoma. In six publications, at least one patient was diagnosed with more than one cancer type. These "2+ diagnoses" include patients with the following cancer types: glioblastoma + non-Hodgkin's Lymphoma + oligodendroglioma (Bakry et al., 2014), glioblastoma + rectal carcinoma (Bakry et al., 2014), glioblastoma + non-Hodgkin's lymphoma (Chmara et al., 2013) or ALL + rectal adenoma (Herkert et al., 2011), acute myeloid leukemia + medulloblastoma (Scott et al., 2007), colon carcinoma + oligodendroglioma (De Rosa et al., 2000), brain tumor + rhabdomyosarcoma (Quesnel et al., 1999).

Tumor Type Standardization

To ensure consistency of names used to describe the tumor types and to reduce the total number of tumor types to consider, we standardized tumor type names, when possible, using parent terms defined in the National Cancer Institute Thesaurus (Sioutos et al., 2007). For example, we grouped B-cell acute lymphoblastic leukemia and T-cell acute lymphoblastic leukemia under their parent term, "acute lymphoblastic leukemia" (ALL).

Variant Pathogenicity Reassessment and Cancer Pathway Association

Because pathogenicity estimation methods varied widely across the studies and because it had been many years since some articles were published, we reassessed pathogenicity using ClinVar and Variant Effect Predictor (VEP) for all studies that provided variant positions (McLaren et al., 2016; Landrum et al., 2018). We used Kyoto Encyclopedia of Genes and Genomes (KEGG) to determine which of the genes, across all studies, are in a known cancer pathway. We used Catalogue of Somatic Mutations in Cancer (COSMIC) to identify genes with known cancer associations (Kanehisa and Goto, 2000; Tate et al., 2019).

RESULTS

Overview of Pediatric Cancer Types and Genes Studied

Researchers have identified *CH* variants across many pediatric cancer types, even though relatively few articles on these topics have been published overall. From 1999 until 2019, an average of ~ 1.2 journal articles per year were published on *CH* variant discovery across a total of 21 cancer types (**Supplementary Data 1: T1**); the highest number ($n = 7$) were published in 2018. The cancer types studied most frequently were ALL (5 publications) (Valentine et al., 2014; Spinella et al., 2015; Moriyama et al., 2017; Diets et al., 2018; Sharapova et al., 2018), non-Hodgkin's Lymphoma (4 publications) (Østergaard et al., 2005; Peters et al., 2009; Bakry et al., 2014; Diets et al., 2018), and medulloblastoma (4 publications) (**Figure 3**; Svojgr et al., 2016; Gröbner et al., 2018; Waszak et al., 2018; Schieffer et al., 2019). ALL and non-Hodgkin's lymphoma are both blood-based cancers, while medulloblastoma is a type of brain tumor (Sandlund et al., 1996; Hunger and Mullighan, 2015; Kumar et al., 2015). In addition, six publications described a patient having been diagnosed with two or more cancer types, which include: glioblastoma + non-Hodgkin's Lymphoma + oligodendroglioma (Bakry et al., 2014), glioblastoma + rectal carcinoma (Bakry et al., 2014), glioblastoma + non-Hodgkin's lymphoma (Chmara

TABLE 1 | Genes with identified *CH* variants and the study(-ies) that identified them.

Gene	Study(-ies)
<i>ANKRD36</i>	Valentine et al. (AML: 13/13)
<i>ATM</i>	Sharapova et al. (ALL: 1/1); Piane et al. (astrocytoma: 1/1); Zhang et al. (HGG: 1/99)
<i>BRCA2</i>	Waszak et al. (MB: 4/1022); Gröbner et al. (MB: 1/42); Svojr et al. (WT: 1/1, MB: 1/1)
<i>CEP55</i>	Spinella et al. (ALL: 2/2)
<i>DDX41</i>	Diness et al. (myeloid neoplasm: 1/1)
<i>DNAH2</i>	Spinella et al. (ALL: 2/2)
<i>ENOSF1</i>	Zhang et al. (HB: 2/2)
<i>FAM83H</i>	Diets et al. (ALL: 1/13)
<i>FLG</i>	Valentine et al. (AML: 2/13)
<i>GRIK1</i>	Diets et al. (ALL: 1/13)
<i>H1-2</i>	Diets et al. (NHL: 1/3)
<i>IRF5</i>	Diets et al. (WT: 1/1)
<i>KMT2C</i>	Valentine et al. (ALL: 6/12; AML: 13/13)
<i>MSH6</i>	Gröbner et al. (HGG: 2/67); Bakry et al. (NHL: 1/5; GB: 1/8); Peters et al. (NHL: 1/1); Scott et al. (MB + AML: 1/1); Okkels et al. (colorectal: 1/1); Østergaard et al. (NHL: 1/1; GB: 1/1);
<i>MUC4</i>	Zhang et al. (HB: 2/2)
<i>MYO18B</i>	Schieffer et al. (MB: 1/1)
<i>NUDT15</i>	Moriyama et al. (ALL: 1/5)
<i>PDE4DIP</i>	Spinella et al. (ALL: 2/2)
<i>PMS2</i>	Gröbner et al. (HGG: 1/67); Bakry et al. (NHL: 1/5; NHL + GB + oligodendroglioma: 1/1; rectal cancer + GB: 1/1); Chmara et al. (GB + NHL: 1/1); Herkert et al. (colon adenoma: 1/1; ALL + rectal adenoma: 1/1); Leenen et al. (NN: 1/1; astrocytoma: 1/1); De Rosa et al. (oligodendroglioma + colon carcinoma: 1/1)
<i>RBMX</i>	Valentine et al. (ALL: 6/12)
<i>RECQL4</i>	Maciaszek et al. (OS: 1/1); Salih et al. (OS: 1/1)
<i>RYR1</i>	Valentine et al. (AML: 9/13)
<i>SDHB</i>	Majumdar et al. (paraganglioma: 1/1)
<i>SLX4</i>	Spinella et al. (ALL: 2/2)
<i>TP53</i>	Quesnel et al. (rhabdomyosarcoma + brain tumor: 1/1)

The details in parentheses next to the author name indicate the cancer type(s) associated with each *CH* variant, the number of patients evaluated in the study, and the number of patients who had a *CH* variant in that gene. For example, Valentine et al. evaluated 13 AML patients and identified a *CH* variant in all 13 patients for the *ANKRD36* gene. AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; HGG, high-grade glioma; MB, medulloblastoma; WT, Wilms tumor; HB, hepatoblastoma; NHL, non-Hodgkin's lymphoma; GB, glioblastoma; NN, neuroepithelial neoplasm; OS, osteosarcoma.

et al., 2013), ALL + rectal adenoma (Herkert et al., 2011), acute myeloid leukemia (AML) + medulloblastoma (Scott et al., 2007), colon carcinoma + oligodendroglioma (De Rosa et al., 2000), and brain tumor + rhabdomyosarcoma (Figure 3; Quesnel et al., 1999).

Sample sizes were limited in most studies (Table 1). Only ~19% of the studies had a sample size >10 for one or more cancer types (Valentine et al., 2014; Zhang et al., 2015; Diets et al., 2018; Gröbner et al., 2018; Waszak et al., 2018). Across all studies, cancer types with more than 10 samples included AML, ALL, high-grade glioma, medulloblastoma (Valentine et al., 2014; Zhang et al., 2015; Diets et al., 2018; Gröbner et al., 2018). Thirteen studies evaluated a single patient (Quesnel et al., 1999;

De Rosa et al., 2000; Okkels et al., 2006; Scott et al., 2007; Peters et al., 2009; Majumdar et al., 2010; Chmara et al., 2013; Piane et al., 2016; Diness et al., 2018; Salih et al., 2018; Sharapova et al., 2018; Maciaszek et al., 2019; Schieffer et al., 2019).

Across all studies, at least one *CH* variant was identified in 25 genes (Table 1). Of these genes, 7 are known to be associated with hereditary cancer predisposition (Table 2). The remaining 18 genes may provide clues about alternative mechanisms of cancer predisposition. For example, one study identified 6 ALL and 13 AML patients with a *CH* variant in *KMT2C* (Valentine et al., 2014), and one study identified two hepatoblastoma patients with a *CH* variant in *MUC4* (Zhang et al., 2018). DNA alterations in *KMT2C* and *MUC4* have been observed in the somatic tissue of medulloblastoma and head and neck squamous cell carcinoma patients, respectively (Tate et al., 2019). Therefore, future studies may reveal that germline mutations in these genes also contribute to pediatric cancer predisposition.

Overview of Studies that Identified *CH* Variants in Genes Associated With Pediatric Cancers

Here we provide an overview of *CH* variant findings specific to genes that have a known association to cancer predisposition. Of the 25 characterized genes identified across all studies (Table 1), COSMIC classifies 7 as being associated with hereditary predisposition for at least one type of pediatric cancer: *ATM*, *BRCA2*, *MSH6*, *PMS2*, *RECQL4*, *SDHB*, and *TP53* (Table 2). Below we provide insight about the functions of these genes, prior associations that have been made for non-compound germline variants, and *CH* variants in these genes.

ATM, *BRCA2*, and *SDHB* are known tumor suppressor genes. Non-compound germline variations in these genes have been associated with leukemias, lymphomas, medulloblastomas, and gliomas (Table 2). Similarly, *CH* variants in *ATM* have been observed in ALL, astrocytoma, and high-grade glioma (Zhang et al., 2015; Piane et al., 2016; Sharapova et al., 2018). *BRCA2* plays important roles in DNA repair, and germline variations in this gene have been associated with breast cancer, ovarian cancer, pancreatic cancer, and leukemia risk (Table 2; Kanehisa and Goto, 2000; Tate et al., 2019). To date, *CH* variants in *BRCA2* have been observed in medulloblastomas and Wilms tumors (Svojr et al., 2016; Gröbner et al., 2018; Waszak et al., 2018). Variations in *SDHB* have been associated with paraganglioma and pheochromocytoma (Table 2); one patient with paraganglioma had a *CH* variant in this gene (Majumdar et al., 2010).

MSH6 and *PMS2* are both considered to be tumor-suppressor and mismatch repair (MMR) genes (Ripperger and Schlegelberger, 2016; Tabori et al., 2017; Tate et al., 2019). Patients with biallelic germline mutations in an MMR gene (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) are considered to have a syndrome known as constitutional mismatch repair disease (CMMRD) (Ripperger and Schlegelberger, 2016; Tabori et al., 2017). Germline mutations in MMR genes have been associated with a predisposition to many different types of cancer (Table 2; Tabori et al., 2017), including hematological malignancies, brain tumors, and digestive tract cancers (Ripperger and Schlegelberger,

TABLE 2 | Details of genes with identified *CH* variants.

Gene	Gene CDS length	Total gene length	Cancer pathway (KEGG)	Germline tumor types (COSMIC)	Role in cancer (COSMIC)
<i>ANKRD36</i>	5,678	151,216	–	–	–
<i>ATM</i>	9,272	146,618	–	Leukemia; lymphoma; medulloblastoma; glioma	TSG
<i>BRCA2</i>	10,415	85,180	Yes	Breast; ovarian; pancreatic; leukemia	TSG
<i>CEP55</i>	1,454	32,480	–	–	–
<i>DDX41</i>	1,802	5,384	–	–	–
<i>DNAH2</i>	13,320	116,390	–	–	–
<i>ENOSF1</i>	1,337	42,344	–	–	–
<i>FAM83H</i>	5,192	14,301	–	–	–
<i>FLG</i>	12,181	23,074	–	–	–
<i>GRIK1</i>	2,887	403,100	–	–	–
<i>H1-2</i>	638	641	–	–	–
<i>IRF5</i>	1,493	12,581	–	–	–
<i>KMT2C</i>	15,426	301,083	–	–	TSG
<i>MSH6</i>	3,591	114,571	Yes	Colorectal; endometrial; ovarian	TSG
<i>MUC4</i>	10,911	65,208	–	–	Oncogene
<i>MYO18B</i>	7,889	288,897	–	–	–
<i>NUDT15</i>	489	9,495	–	–	–
<i>PDE4DIP</i>	7,104	240,105	–	–	Fusion
<i>PMS2</i>	2,463	38,181	–	Colorectal; endometrial; ovarian; medulloblastoma; glioma	TSG
<i>RBMX</i>	1,166	32,760	–	–	–
<i>RECQL4</i>	3,252	6,557	–	Osteosarcoma; skin basal cell; skin squamous cell	Oncogene; TSG
<i>RYR1</i>	14,835	153,873	–	–	–
<i>SDHB</i>	778	35,310	–	Paraganglioma; pheochromocytoma	TSG
<i>SLX4</i>	5,488	30,425	–	–	–
<i>TP53</i>	1,194	25,771	Yes	Breast; sarcoma; adrenocortical carcinoma; glioma; multiple other tumor types	Oncogene; TSG; fusion

The average CDS gene length across all studies was 5,610 and the average total gene length was 95,022. *ATM*, *BRCA2*, *MSH6*, *PMS2*, *RECQL4*, *SDHB*, and *TP53* are all associated with germline tumor types and are known tumor suppressor genes as classified by COSMIC. *BRCA2*, *MSH6*, and *TP53* are all part of a cancer pathway as classified by KEGG. A value of “–” indicates that an association was not observed. TSG, tumor suppressor gene.

2016). To date, *CH* variants in *MSH6* have been identified in patients with high-grade glioma, non-Hodgkin’s lymphoma, glioblastoma, medulloblastoma, AML, or colorectal cancer (Østergaard et al., 2005; Okkels et al., 2006; Scott et al., 2007; Peters et al., 2009; Bakry et al., 2014; Gröbner et al., 2018). *PMS2* *CH* variants have been identified in high-grade glioma, non-Hodgkin’s lymphoma, glioblastoma, oligodendroglioma, rectal cancer, colon adenoma, ALL, recta adenoma, neuroepithelial neoplasm, astrocytoma, and colon carcinoma (De Rosa et al., 2000; Herkert et al., 2011; Leenen et al., 2011; Chmara et al., 2013; Bakry et al., 2014; Gröbner et al., 2018).

RECQL4 and *TP53* both have more than one role in cancer (Tate et al., 2019). *RECQL4* is involved in many intracellular regulatory pathways and can act either as an oncogene or a tumor suppressor gene (Kellermayer, 2006; Arora et al., 2016). Germline cancer associations for *RECQL4* include osteosarcoma, skin basal cell, and skin squamous cell (Table 2). *CH* variants in *RECQL4* have been associated with osteosarcoma in two studies (Salih et al., 2018; Maciaszek et al., 2019). *TP53* is involved in many cancer pathways, and germline variation in this gene has been associated with a wide range of tumor types (Table 2). One patient with

rhabdomyosarcoma + brain tumor had a *CH* variant in *TP53* (Quesnel et al., 1999).

Methodologies Used to Identify *CH* Variants and Assess Pathogenicity

The sequencing methods used were highly variable across the studies (Table 3 and Supplementary Data 1:T1). Of the studies in this review, 10 indicated that they used one or more forms of NGS sequencing (whole-exome, whole-genome, RNA-seq) and thus were able to sample *CH* variants broadly (Valentine et al., 2014; Spinella et al., 2015; Zhang et al., 2015, 2018; Diets et al., 2018; Diness et al., 2018; Gröbner et al., 2018; Waszak et al., 2018; Maciaszek et al., 2019; Schieffer et al., 2019). Thirteen studies indicated that they used non-NGS methods (Sanger, Direct sequencing, SNP-array, Multiplex ligation-dependent probe amplification) and were therefore limited in the number of genes and *CH* variants analyzed (Quesnel et al., 1999; De Rosa et al., 2000; Østergaard et al., 2005; Okkels et al., 2006; Scott et al., 2007; Herkert et al., 2011; Leenen et al., 2011; Chmara et al., 2013; Bakry et al., 2014; Piane et al., 2016; Svojgr et al., 2016; Moriyama et al., 2017; Sharapova et al., 2018). Three clinical/case report studies did not describe the exact DNA sequencing technology

TABLE 3 | Methods used by each study for identification and evaluation of *CH* variants.

References	Sequencing technology	How <i>CH</i> variant identified	Post sequencing pathogenicity evaluation criteria
Maciaszek et al. (2019)	WGS	Mendelian inheritance	ACMG/AMP guidelines
Schieffer et al. (2019)	WES; Sanger	Mendelian inheritance	ACMG/AMP guidelines
Zhang et al. (2018)	WGS	Mendelian inheritance	<i>SIFT</i> ; <i>PolyPhen-2</i> ; <i>MutationTaster</i> (Schwarz et al., 2014), M-CAP; and AMP/ACMG guidelines
Diness et al. (2018)	WES	Mendelian inheritance	<i>CADD</i> (Kircher et al., 2014); gene expression analysis
Sharapova et al. (2018)	Sanger	Mendelian inheritance	<i>PolyPhen-2</i>
Waszak et al. (2018)	WGS; WES; RNA-seq	Multiple sites within the same gene were phased with paired-end RNA sequencing data and individual sites were merged to calculate haplotype-specific expression ratios.	ClinVar
Diets et al. (2018)	WES	Mendelian inheritance	<i>SIFT</i> ; <i>PolyPhen2</i> ; <i>CADD</i>
Gröbner et al. (2018)	WGS; WES	Used Platypus which is a haplotype-based variant caller (Rimmer et al., 2014)	<i>CADD</i> ; <i>Mutation Assessor</i> (Reva et al., 2011)
Salih et al. (2018)	Sanger (inferred)	Mendelian inheritance	–
Moriyama et al. (2017)	Sanger	PHASE was used to infer haplotypes (Stephens and Scheet, 2005)	–
Svojgr et al. (2016)	SNP-array	Mendelian inheritance	–
Piane et al. (2016)	Sanger	Mendelian inheritance	<i>SIFT</i> ; <i>Polyphen</i> ; <i>MutationTaster</i>
Spinella et al. (2015)	WES	Mendelian inheritance	<i>SIFT</i> ; <i>PolyPhen-2</i> ; and hidden Markov models
Zhang et al. (2015)	WGS; WES; RNA-seq	Used RNA-seq data to determine <i>CH</i> nature of variant	ACMG/AMP guidelines; genetic database; medical literature; computational predictions; and second hits identified in the tumor genome
Valentine et al. (2014)	WES	Mendelian inheritance	Filtered for functional consequences (e.g., non-synonymous and coding)
Bakry et al. (2014)	Sanger	Mendelian inheritance	Algorithms to predict RNA/protein disruption
Chmara et al. (2013)	Direct sequencing; MLPA	Mendelian inheritance	–
Herkert et al. (2011)	Direct sequencing; MLPA	Mendelian inheritance	<i>SIFT</i> ; <i>AlignGVGD</i> (Tavtigian et al., 2008); <i>PolyPhen-2</i> ; and RNA splice site prediction programs
Leenen et al. (2011)	Sanger	Mendelian inheritance	Literature search
Majumdar et al. (2010)	Sanger (inferred)	Mendelian inheritance	–
Peters et al. (2009)	Sanger (inferred)	Mendelian inheritance	–
Scott et al. (2007)	Sanger	Mendelian inheritance	Literature search
Okkels et al. (2006)	Sanger	Mendelian inheritance	–
Østergaard et al. (2005)	Sanger	Mendelian inheritance	–
De Rosa et al. (2000)	Sanger	Mendelian inheritance	–
Quesnel et al. (1999)	Sanger	Mendelian inheritance	mRNA assay

A value of “–” indicates that no clear description was provided by the authors. WGS, whole genome sequencing; WES, whole exome sequencing; MLPA, Multiplex Ligation-dependent Probe Amplification.

used (Peters et al., 2009; Majumdar et al., 2010; Salih et al., 2018); however, we inferred that Sanger sequencing was inferred in these three studies because single genes were the focus of the studies, parent DNA was also sequenced, and DNA variant locations were provided by the authors.

Across all studies, the methodology used to estimate haplotypes varied considerably (Table 3 and Supplementary Data 1:T1). Four of the studies used a computer-based phasing algorithm to estimate haplotypes (Zhang et al., 2015; Moriyama et al., 2017; Gröbner et al., 2018; Waszak et al., 2018). One study used evidence from RNA-seq data that the alternate alleles were on different chromosomes (Zhang et al., 2015). All other studies that performed phasing

used sequence data from the patient and his/her parent(s) (i.e., Mendelian inheritance; Quesnel et al., 1999; De Rosa et al., 2000; Okkels et al., 2006; Scott et al., 2007; Peters et al., 2009; Majumdar et al., 2010; Herkert et al., 2011; Leenen et al., 2011; Chmara et al., 2013; Bakry et al., 2014; Valentine et al., 2014; Spinella et al., 2015; Piane et al., 2016; Svojgr et al., 2016; Diets et al., 2018; Diness et al., 2018; Salih et al., 2018; Sharapova et al., 2018; Zhang et al., 2018; Maciaszek et al., 2019; Schieffer et al., 2019).

The methods used to classify variant pathogenicity also varied widely from study to study (Table 3 and Supplementary Data 1:T1). Eight studies used predictive algorithms as the sole means of pathogenicity assessment

(Herkert et al., 2011; Bakry et al., 2014; Spinella et al., 2015; Piane et al., 2016; Diets et al., 2018; Gröbner et al., 2018; Sharapova et al., 2018; Waszak et al., 2018). Four studies used guidelines set forth by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) (Zhang et al., 2015, 2018; Maciaszek et al., 2019; Schieffer et al., 2019). Two studies used literature searches to look for known effects of the identified variants (Scott et al., 2007; Leenen et al., 2011). Two studies analyzed RNA expression of altered alleles (Quesnel et al., 1999; Diness et al., 2018). One study did not directly assess pathogenicity; rather they reported on variants that met specific criteria such as the variant being a non-synonymous substitution (Valentine et al., 2014). Nine studies made no mention of assessing CH variant pathogenicity (De Rosa et al., 2000; Østergaard et al., 2005; Okkels et al., 2006; Peters et al., 2009; Majumdar et al., 2010; Chmara et al., 2013; Svojr et al., 2016; Moriyama et al., 2017; Salih et al., 2018).

There were a total of 18 CH variants where both alleles were reported as pathogenic or likely pathogenic by the authors of the study (**Supplementary Data 1:T2**). Because some of the studies were conducted years ago and pathogenicity classifications may have been updated, we re-assessed the pathogenicity of all variants that were provided by the authors of the studies. Using ClinVar and VEP (as SIFT and PolyPhen scores), we were able to confirm pathogenicity for 5 of the 18 CH variants with one or more of our reassessment methods (**Supplementary Data 1:T2**). These confirmed variants were in *MYO18B* (Schieffer et al., 2019), *BRCA2* (Waszak et al., 2018), *FAM83H* (Diets et al., 2018), and *HIST1H1C* (Diets et al., 2018). In addition, we were able to classify 3 CH variants as pathogenic or likely pathogenic that were not indicated as such in the original study. These variants were in *RECQL4* (Salih et al., 2018), *PMS2* (Herkert et al., 2011), and *MSH6* (Scott et al., 2007). For the remaining variants, either only one allele from a CH pair was able to be classified, or no information was available in the databases (**Supplementary Data 1:T2**).

DISCUSSION

Observations and Future Directions

Research to date has highlighted that CH variants are observed relatively rarely but occur in many different genes and a diverse array of pediatric tumor types. However, our knowledge of the roles that CH variants play in pediatric cancers is only in its infancy. Prior studies have focused primarily on candidate genes, have been limited to individual cancer types, or have been limited by small sample sizes (**Table 1**). Due to these issues, important CH variants may have been missed. For example, in AML, tumor suppressor genes that are known to harbor germline risk variants include *BRIP1*, *FANCA*, *FANCC*, *FANCD2*, *FANCE*, *FANCF*, *FANCG*, *PALB2*, and *SBDS* (Sondka et al., 2018; Tate et al., 2019). However, no study to date has found CH variants in any of these genes for AML patients. Similarly, *ATM*, *NBN*, *PALB2*, *PMS2*, *PTCH1*, and *SUFU* are tumor suppressor and germline risk genes for medulloblastoma, but CH variants in these genes have not been found in prior studies.

One factor that may have contributed to a lack of identifying CH variants in known cancer-predisposition genes may be filtering based on minor allele frequencies (MAF). Commonly, individual alleles with a MAF >1% (or >0.5%) in a control population are excluded from analyses because they are assumed to have benign effects (Richards et al., 2015; Niroula and Vihinen, 2019). Yet, this traditional filtering criterion may be too stringent for identifying pathogenic CH variants. By definition, two pathogenic alleles must be present for a recessive phenotype to manifest itself. Suppose, for example, that one allele in a given gene is relatively rare, having been observed in 0.8% of the population. Now suppose that a second allele is present at a different locus within the same gene and that this allele has a population prevalence of 5%. MAF filtering considers each locus separately, so the latter variant would be excluded by the 1% threshold, causing this potentially pathogenic CH variant to be overlooked. Assuming non-consanguinity and using the probability multiplication rule, the population prevalence of this particular CH variant would be estimated at approximately 0.04% (5% x 0.8%). Although care must be taken to consider other possible combinations of *in trans* alleles in this gene (Eilbeck et al., 2017), this example illustrates that traditional MAF filtering may be too simplistic for CH variant analysis. Furthermore, researchers must account for any *homozygous*, non-reference genotypes that have been observed for either allele in healthy individuals (Kamphans et al., 2013).

Further complicating matters, it is difficult to estimate the *a priori* expectation of finding a CH variant in a given gene. The longer a gene, the higher the probability that two pathogenic alleles will occur within that gene. Accordingly, genome-wide studies may be biased toward identifying CH variants in longer genes. For example, across the 26 studies covered in this review, the average coding sequence (CDS) length for genes with an identified CH variant was 5,610 bases (median: 3,591 bases) (**Table 2**). Comparatively, using data from GENCODE, we calculated the average CDS length of protein coding genes across the human genome to be ~1,796 bases (median: 1,338) (Frankish et al., 2019). Efforts should be made to help alleviate this bias and other confounding factors. For example, Itan et al. developed the Gene Damage Index to prioritize genes based on CDS length, protein complexity, paralog count, and evolutionary pressures (Itan et al., 2015). When prioritizing CH variants at the gene level in this way, the number of false-positive genes may be reduced.

Available Pediatric Cancer Data

The future is primed for more rapid discovery of genetic factors involved in pediatric cancers and a clearer understanding of the roles that CH variants play in pediatric cancer development and progression. Publicly available data are becoming readily accessible for researchers to study, and more data will become available over the next few years. For example, the NIH-funded Gabriella Miller Kids First (GMKF) initiative is sequencing germline DNA from hundreds of trios (affected child and both parents). This initiative is enabling researchers to obtain Illumina-based, whole-genome sequencing (WGS) data for these trios across many childhood cancer types (and other pediatric diseases) (Gabriella Miller Kids First

Pediatric Research Program²). Currently, the GMKF repository contains sequencing data for Ewing sarcoma (250 trios) and neuroblastoma (470 trios). In the coming years, the GMKF repository is expected to include data from pediatric cancers such as osteosarcoma, leukemia, enchondromatosis, brain tumors, myeloid malignancies, and others. This repository will prove valuable as parental data may allow for better identification of *CH* variants and *de novo* mutations (Francioli et al., 2017; Choi et al., 2018).

Frequently it is infeasible to obtain sequencing data from a proband's parents due to cost limitations or logistical challenges (a parent may not consent to participate or may be unavailable, a parent may be deceased, the child may be adopted, etc.). Projects such as Therapeutically Applicable Research to Generate Effective Treatments (TARGET), St. Jude Cloud, and the Children's Brain Tumor Tissue Consortium (CBTTC) have sequenced DNA for thousands of patients across many pediatric diseases, but sequencing data are available for the proband only in these resources (Children's Hospital of Philadelphia³; Downing et al., 2012; Office of Cancer Genomics, 2013). At the time of this writing, TARGET included matched tumor-normal, multi-omic data across six types of pediatric cancer, including AML ($n \approx 50$), neuroblastoma ($n \approx 228$), Wilms tumor ($n \approx 81$), osteosarcoma ($n \approx 89$), clear cell sarcoma of the kidney ($n \approx 13$), and rhabdoid tumor ($n \approx 43$) (Office of Cancer Genomics, 2013). The St. Jude Cloud contained matched tumor-normal data for ~ 42 pediatric cancer types/subtypes, which encompassed $\sim 2,168$ patients (Downing et al., 2012). Some of the datasets with the highest number of patients in the St. Jude Cloud included ALL ($n \approx 260$), AML ($n \approx 189$), HGG ($n \approx 163$), and neuroblastoma ($n \approx 135$) (Downing et al., 2012). Eight cancer types are represented by over 100 patients, while 14 cancer types had data for 10 or fewer patients. The CBTCC is a collaborative effort dedicated to the study and treatment of pediatric brain tumors (Children's Hospital of Philadelphia). At the time of this writing, CBTCC included genomic data for ~ 871 patients across ~ 38 pediatric brain tumor types. Although these databases do not contain parental genome data, computer-based algorithms can often determine a variant's parent-of-origin using haplotype reference panels (Browning and Browning, 2011).

²Gabriella Miller Kids First Pediatric Research Program National Institutes of Health Office of Strategic Coordination - The Common Fund. Available online at: <https://commonfund.nih.gov/kidsfirst> (accessed March 26, 2019).

³Children's Hospital of Philadelphia Children's Brain Tumor Tissue Consortium. *Children's Brain Tumor Tissue Consortium*. Available online at: <https://cbttc.org/> (accessed January 31, 2020).

REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248
- Arora, A., Agarwal, D., Abdel-Fatah, T. M., Lu, H., Croteau, D. L., Moseley, P., et al. (2016). RECQL4 helicase has oncogenic potential in sporadic breast cancers. *J. Pathol.* 238, 495–501. doi: 10.1002/path.4681

Thanks to these public and non-profit efforts, we have entered an era in which researchers can shed more light on the genes and pathways that influence specific pediatric cancers through the use of genome-wide, publicly available data. We advocate that researchers take advantage of these resources.

CONCLUSION

Many discoveries about *CH* variants and their association with pediatric cancers have been made over the last 20 years; the role of *CH* variants in cancer and developmental pathways and the prevalence of these variants in pediatric cancers are beginning to be uncovered. Through the works discussed in this review, much insight has been gained. As future studies are conducted on *CH* variants in pediatric cancers, we anticipate that *CH* variants will play a more prominent role in elucidating disease mechanisms. This heightened knowledge could expand this field of study and eventually lead to the development of targeted treatments. Furthermore, having an understanding of *CH* variants and their role in disease development could prove beneficial for disease monitoring. For example, a child with a malignancy who has a germline risk variant in a known predisposition gene and/or key pathway could be more regularly monitored and assessed for additional tumor development (Milanese and Wang, 2019). If risk variants are in genes associated with one particular type of cancer, screening efforts could be more directed than general cancer screening. More specific screening could allow for earlier detection of tumors, sooner treatments, and prophylactic measures. Finally, a heightened knowledge of *CH* variants could lead to an expansion of our understanding of other pediatric diseases, such as birth defects, and even inherited disorders that arise in adults.

AUTHOR CONTRIBUTIONS

DM wrote the manuscript and performed critical analysis. SP provided edits and critical feedback.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00493/full#supplementary-material>

- Bakry, D., Aronson, M., Durno, C., Rimawi, H., Farah, R., Alharbi, Q. K., et al. (2014). Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: report from the constitutional mismatch repair deficiency consortium. *Eur. J. Cancer* 50, 987–996. doi: 10.1016/j.ejca.2013.12.005
- Browning, S. R., and Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81, 1084–1097. doi: 10.1086/521987

- Browning, S. R., and Browning, B. L. (2011). Haplotype phasing: existing methods and new developments. *Nat. Rev. Genet.* 12, 703–714. doi: 10.1038/nrg3054
- Chmara, M., Wernstedt, A., Wasag, B., Peeters, H., Renard, M., Beert, E., et al. (2013). Multiple pilomatricomas with somatic CTNBN1 mutations in children with constitutive mismatch repair deficiency. *Genes Chromosomes Cancer* 52, 656–664. doi: 10.1002/gcc.22061
- Choi, Y., Chan, A. P., Kirkness, E., Telenti, A., and Schork, N. J. (2018). Comparison of phasing strategies for whole human genomes. *PLoS Genet.* 14:e1007308. doi: 10.1371/journal.pgen.1007308
- De Rosa, M., Fasano, C., Panariello, L., Scarano, M. I., Belli, G., Iannelli, A., et al. (2000). Evidence for a recessive inheritance of Turcot's syndrome caused by compound heterozygous mutations within the PMS2 gene. *Oncogene* 19, 1719–1723. doi: 10.1038/sj.onc.1203447
- Dean, S. J., and Farmer, M. (2017). Pediatric cancer genetics. *Curr. Opin. Pediatr.* 29, 629–633. doi: 10.1097/MOP.0000000000000543
- Delaneau, O., Zagury, J.-F., and Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* 10, 5–6. doi: 10.1038/nmeth.2307
- Diets, I. J., Waanders, E., Ligtenberg, M. J., van Bladel, D. A. G., Kamping, E. J., Hoogerbrugge, P. M., et al. (2018). High yield of pathogenic germline mutations causative or likely causative of the cancer phenotype in selected children with cancer. *Clin. Cancer Res.* 24, 1594–1603. doi: 10.1158/1078-0432.CCR-17-1725
- Diness, B. R., Risom, L., Frandsen, T. L., Hansen, B., Andersen, M. K., Schmiegelow, K., et al. (2018). Putative new childhood leukemia cancer predisposition syndrome caused by germline bi-allelic missense mutations in DDX41. *Genes Chromosomes Cancer* 57, 670–674. doi: 10.1002/gcc.22680
- Downing, J. R., Wilson, R. K., Zhang, J., Mardis, E. R., Pui, C.-H., Ding, L., et al. (2012). The pediatric cancer genome project. *Nat. Genet.* 44, 619–622. doi: 10.1038/ng.2287
- Eilbeck, K., Quinlan, A., and Yandell, M. (2017). Settling the score: variant prioritization and Mendelian disease. *Nat. Rev. Genet.* 18, 599–612. doi: 10.1038/nrg.2017.52
- Francioli, L. C., Cretu-Stancu, M., Garimella, K. V., Fromer, M., Kloosterman, W. P., Genome of the Netherlands consortium, et al. (2017). A framework for the detection of *de novo* mutations in family-based sequencing data. *Eur. J. Hum. Genet.* 25, 227–233. doi: 10.1038/ejhg.2016.147
- Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* 47, D766–D773. doi: 10.1093/nar/gky955
- Gröbner, S. N., Worst, B. C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V. A., et al. (2018). The landscape of genomic alterations across childhood cancers. *Nature* 555, 321–327. doi: 10.1038/nature25480
- Guo, X. E., Ngo, B., Modrek, A. S., and Lee, W.-H. (2014). Targeting tumor suppressor networks for cancer therapeutics. *Curr. Drug Targets* 15, 2–16. doi: 10.2174/1389450114666140106095151
- Herkert, J. C., Niessen, R. C., Olderode-Berends, M. J. W., Veenstra-Knol, H. E., Vos, Y. J., van der Klift, H. M., et al. (2011). Paediatric intestinal cancer and polyposis due to bi-allelic PMS2 mutations: case series, review and follow-up guidelines. *Eur. J. Cancer* 47, 965–982. doi: 10.1016/j.ejca.2011.01.013
- Hunger, S. P., and Mullighan, C. G. (2015). Acute lymphoblastic leukemia in children. *N. Engl. J. Med.* 373, 1541–1552. doi: 10.1056/NEJMra1400972
- Itan, Y., Shang, L., Boisson, B., Patin, E., Bolze, A., Moncada-Vélez, M., et al. (2015). The human gene damage index as a gene-level approach to prioritizing exome variants. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13615–13620. doi: 10.1073/pnas.1518646112
- Kamphans, T., Sabri, P., Zhu, N., Heinrich, V., Mundlos, S., Robinson, P. N., et al. (2013). Filtering for compound heterozygous sequence variants in non-consanguineous pedigrees. *PLoS ONE* 8:e70151. doi: 10.1371/journal.pone.0070151
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kellermayer, R. (2006). The versatile RECQL4. *Genet. Med.* 8, 213–216. doi: 10.1097/01.gim.0000214457.58378.1a
- Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. doi: 10.1038/ng.2892
- Kumar, L. P., Deepa, S. F. A. J., Moince, I., Suresh, P., and Naidu, K. V. J. R. (2015). Medulloblastoma: a common pediatric tumor: prognostic factors and predictors of outcome. *Asian J. Neurosurg.* 10:50. doi: 10.4103/1793-5482.151516
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 46, D1062–D1067. doi: 10.1093/nar/gkx1153
- Leenen, C. H. M., Geurts-Giele, W. R. R., Dubbink, H. J., Reddingius, R., van den Ouweland, A. M., Tops, C. M. J., et al. (2011). Pitfalls in molecular analysis for mismatch repair deficiency in a family with biallelic pms2 germline mutations. *Clin. Genet.* 80, 558–565. doi: 10.1111/j.1399-0004.2010.01608.x
- Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., et al. (2016). Reference-based phasing using the haplotype reference consortium panel. *Nat. Genet.* 48, 1443–1448. doi: 10.1038/ng.3679
- Maciaszek, J. L., Oak, N., Chen, W., Hamilton, K. V., McGee, R. B., Nuccio, R., et al. (2019). Enrichment of heterozygous germline RECQL4 loss-of-function variants in pediatric osteosarcoma. *Cold Spring Harb. Mol. Case Stud.* 5:a004218. doi: 10.1101/mcs.a004218
- Majumdar, S., Friedrich, C. A., Koch, C. A., Megason, G. C., Fratkin, J. D., and Moll, G. W. (2010). Compound heterozygous mutation with a novel splice donor region DNA sequence variant in the succinate dehydrogenase subunit B gene in malignant paraganglioma. *Pediatr. Blood Cancer* 54, 473–475. doi: 10.1002/psc.22338
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biol.* 17:122. doi: 10.1186/s13059-016-0974-4
- Milanesi, J.-S., and Wang, E. (2019). Germline mutations and their clinical applications in cancer. *Breast Cancer Manage.* 8:BMT23. doi: 10.2217/bmt-2019-0005
- Moriyama, T., Yang, Y.-L., Nishii, R., Ariffin, H., Liu, C., Lin, T.-N., et al. (2017). Novel variants in NUDT15 and thiopurine intolerance in children with acute lymphoblastic leukemia from diverse ancestry. *Blood* 130, 1209–1212. doi: 10.1182/blood-2017-05-782383
- Nafisinia, M., Riley, L. G., Gold, W. A., Bhattacharya, K., Broderick, C. R., Thorburn, D. R., et al. (2017). Compound heterozygous mutations in glycyl-tRNA synthetase (GARS) cause mitochondrial respiratory chain dysfunction. *PLoS ONE* 12:e0178125. doi: 10.1371/journal.pone.0178125
- Ng, P. C., and Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814. doi: 10.1093/nar/gkg509
- Niroula, A., and Vihinen, M. (2019). How good are pathogenicity predictors in detecting benign variants? *PLoS Comput. Biol.* 15:e1006481. doi: 10.1371/journal.pcbi.1006481
- Office of Cancer Genomics (2013). *Therapeutically Applicable Research to Generate Effective Treatments*. Office of Cancer Genomics. Available online at: <https://ocg.cancer.gov/programs/target> (accessed January 31, 2020).
- Okkels, H., Sunde, L., Lindorff-Larsen, K., Thorlacius-Ussing, O., Gandrup, P., Lindebjerg, J., et al. (2006). Polyposis and early cancer in a patient with low penetrant mutations in MSH6 and APC: hereditary colorectal cancer as a polygenic trait. *Int. J. Colorectal Dis.* 21, 847–850. doi: 10.1007/s00384-006-0086-9
- Østergaard, J. R., Sunde, L., and Okkels, H. (2005). Neurofibromatosis von Recklinghausen type I phenotype and early onset of cancers in siblings compound heterozygous for mutations in MSH6. *Am. J. Med. Genet.* 139A, 96–105. doi: 10.1002/ajmg.a.30998
- Parsons, D. W., Roy, A., Yang, Y., Wang, T., Scollon, S., Bergstrom, K., et al. (2016). Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors. *JAMA Oncol.* 2, 616–624. doi: 10.1001/jamaoncol.2015.5699
- Peters, A., Born, H., Ettinger, R., Levonian, P., and Jedele, K. B. (2009). Compound heterozygosity for MSH6 mutations in a pediatric lymphoma patient. *J. Pediatr. Hematol. Oncol.* 31, 113–115. doi: 10.1097/MPH.0b013e31818b3774
- Phillips, S. M., Padgett, L. S., Leisenring, W. M., Stratton, K. K., Bishop, K., Krull, K. R., et al. (2015). Survivors of childhood cancer in the United States: prevalence and burden of morbidity. *Cancer Epidemiol. Biomarkers Prev.* 24, 653–663. doi: 10.1158/1055-9965.EPI-14-1418

- Piane, M., Molinaro, A., Soresina, A., Costa, S., Maffei, M., Germani, A., et al. (2016). Novel compound heterozygous mutations in a child with Ataxia-Telangiectasia showing unrelated cerebellar disorders. *J. Neurol. Sci.* 371, 48–53. doi: 10.1016/j.jns.2016.10.014
- Quesnel, S., Verselis, S., Portwine, C., Garber, J., White, M., Feunteun, J., et al. (1999). p53 compound heterozygosity in a severely affected child with Li-Fraumeni syndrome. *Oncogene* 18, 3970–3978. doi: 10.1038/sj.onc.1202783
- Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 39:e118. doi: 10.1093/nar/gkr407
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17, 405–424. doi: 10.1038/gim.2015.30
- Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R. F., WGS500 Consortium, et al. (2014). Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* 46, 912–918. doi: 10.1038/ng.3036
- Ripperger, T., and Schlegelberger, B. (2016). Acute lymphoblastic leukemia and lymphoma in the context of constitutional mismatch repair deficiency syndrome. *Eur. J. Med. Genet.* 59, 133–142. doi: 10.1016/j.ejmg.2015.12.014
- Rodenburg, R. J. (2018). The functional genomics laboratory: functional validation of genetic variants. *J. Inherit. Metab. Dis.* 41, 297–307. doi: 10.1007/s10545-018-0146-7
- Salih, A., Inoue, S., and Onwuzurike, N. (2018). Rothmund-Thomson syndrome (RTS) with osteosarcoma due to RECQL4 mutation. *BMJ Case Rep.* 2018:bcr2017222384. doi: 10.1136/bcr-2017-222384
- Sandlund, J. T., Downing, J. R., and Crist, W. M. (1996). Non-Hodgkin's lymphoma in childhood. *N. Engl. J. Med.* 334, 1238–1248. doi: 10.1056/NEJM199605093341906
- Schieffer, K. M., Varga, E., Miller, K. E., Agarwal, V., Koboldt, D. C., Brennan, P., et al. (2019). Expanding the clinical history associated with syndromic Klippel-Feil: a unique case of comorbidity with medulloblastoma. *Eur. J. Med. Genet.* 62:103701. doi: 10.1016/j.ejmg.2019.103701
- Schwarz, J. M., Cooper, D. N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* 11, 361–362. doi: 10.1038/nmeth.2890
- Scott, R. H., Mansour, S., Pritchard-Jones, K., Kumar, D., MacSweeney, F., and Rahman, N. (2007). Medulloblastoma, acute myelocytic leukemia and colonic carcinomas in a child with biallelic MSH6 mutations. *Nat. Clin. Pract. Oncol.* 4, 130–134. doi: 10.1038/ncponc0719
- Sharapova, S. O., Valochnik, A. V., Guryanova, I. E., Sakovich, I. S., and Aleinikova, O. V. (2018). Novel biallelic ATM mutations coexist with a mosaic form of triple X syndrome in an 11-year-old girl at remission after T cell acute leukemia. *Immunogenetics* 70, 613–617. doi: 10.1007/s00251-018-1056-4
- Sioutos, N., de Coronado, S., Haber, M. W., Hartel, F. W., Shaiu, W.-L., and Wright, L. W. (2007). NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J. Biomed. Inform.* 40, 30–43. doi: 10.1016/j.jbi.2006.02.013
- Sondka, Z., Bamford, S., Cole, C. G., Ward, S. A., Dunham, I., and Forbes, S. A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* 18, 696–705. doi: 10.1038/s41568-018-0060-1
- Spinella, J.-F., Healy, J., Saillour, V., Richer, C., Cassart, P., Ouimet, M., et al. (2015). Whole-exome sequencing of a rare case of familial childhood acute lymphoblastic leukemia reveals putative predisposing mutations in Fanconi anemia genes. *BMC Cancer* 15:539. doi: 10.1186/s12885-015-1549-6
- Stephens, M., and Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* 76, 449–462. doi: 10.1086/428594
- Svojr, K., Sumerauer, D., Puchmajerova, A., Vicha, A., Hrusak, O., Michalova, K., et al. (2016). Fanconi anemia with biallelic FANCD1/BRCA2 mutations - case report of a family with three affected children. *Eur. J. Med. Genet.* 59, 152–157. doi: 10.1016/j.ejmg.2015.11.013
- Sweet-Cordero, E. A., and Biegel, J. A. (2019). The genomic landscape of pediatric cancers: Implications for diagnosis and treatment. *Science* 363, 1170–1175. doi: 10.1126/science.aaw3535
- Tabori, U., Hansford, J. R., Achatz, M. I., Kratz, C. P., Plon, S. E., Frebourg, T., et al. (2017). Clinical management and tumor surveillance recommendations of inherited mismatch repair deficiency in childhood. *Clin. Cancer Res.* 23, e32–e37. doi: 10.1158/1078-0432.CCR-17-0574
- Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47, D941–D947. doi: 10.1093/nar/gky1015
- Tavtigian, S. V., Byrnes, G. B., Goldgar, D. E., and Thomas, A. (2008). Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum. Mutat.* 29, 1342–1354. doi: 10.1002/humu.20896
- Tewhey, R., Bansal, V., Torkamani, A., Topol, E. J., and Schork, N. J. (2011). The importance of phase information for human genomics. *Nat. Rev. Genet.* 12, 215–223. doi: 10.1038/nrg2950
- Valentine, M. C., Linabery, A. M., Chasnoff, S., Hughes, A. E. O., Mallaney, C., Sanchez, N., et al. (2014). Excess congenital non-synonymous variation in leukemia-associated genes in MLL- infant leukemia: a Children's Oncology Group report. *Leukemia* 28, 1235–1241. doi: 10.1038/leu.2013.367
- Wang, L.-H., Wu, C.-F., Rajasekaran, N., and Shin, Y. K. (2018). Loss of tumor suppressor gene function in human cancer: an overview. *Cell. Physiol. Biochem.* 51, 2647–2693. doi: 10.1159/000495956
- Waszak, S. M., Northcott, P. A., Buchhalter, I., Robinson, G. W., Sutter, C., Groebner, S., et al. (2018). Spectrum and prevalence of genetic predisposition in medulloblastoma: a retrospective genetic study and prospective validation in a clinical trial cohort. *Lancet Oncol.* 19, 785–798. doi: 10.1016/S1470-2045(18)30242-0
- Weinberg, R. (2013). *The Biology of Cancer, 2nd Edn.* New York, NY: Garland Science.
- Zhang, J., Walsh, M. F., Wu, G., Edmonson, M. N., Gruber, T. A., Easton, J., et al. (2015). Germline mutations in predisposition genes in pediatric cancer. *N. Engl. J. Med.* 373, 2336–2346. doi: 10.1056/NEJMoa1508054
- Zhang, L., Jin, Y., Zheng, K., Wang, H., Yang, S., Lv, C., et al. (2018). Whole-genome sequencing identifies a novel variation of WAS gene coordinating with heterozygous germline mutation of APC to enhance hepatoblastoma oncogenesis. *Front. Genet.* 9:668. doi: 10.3389/fgene.2018.00668
- Zheng, G. X. Y., Lau, B. T., Schnall-Levin, M., Jarosz, M., Bell, J. M., Hindson, C. M., et al. (2016). Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* 34, 303–311. doi: 10.1038/nbt.3432

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

Copyright © 2020 Miller and Piccolo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.