

---

# Germline predisposition to pediatric Ewing sarcoma is characterized by inherited pathogenic variants in DNA damage repair genes

## Authors

Riaz Gillani, Sabrina Y. Camp, Seunghun Han, ...,  
Brian D. Crompton, Saud H. AlDubayan,  
Eliezer M. Van Allen

## Correspondence

[saud\\_aldubayan@dfci.harvard.edu](mailto:saud_aldubayan@dfci.harvard.edu) (S.H.A.),  
[eliezerm\\_vanallen@dfci.harvard.edu](mailto:eliezerm_vanallen@dfci.harvard.edu) (E.M.V.A.)



# Germline predisposition to pediatric Ewing sarcoma is characterized by inherited pathogenic variants in DNA damage repair genes

Riaz Gillani,<sup>1,2,3,4</sup> Sabrina Y. Camp,<sup>2,5</sup> Seunghun Han,<sup>2,6</sup> Jill K. Jones,<sup>6</sup> Hoyin Chu,<sup>2,5</sup> Schuyler O'Brien,<sup>7</sup> Erin L. Young,<sup>7</sup> Lucy Hayes,<sup>7</sup> Gareth Mitchell,<sup>7</sup> Trent Fowler,<sup>7</sup> Alexander Gusev,<sup>5,8</sup> Junne Kamihara,<sup>1,3,4</sup> Katherine A. Janeway,<sup>1,2,3,4</sup> Joshua D. Schiffman,<sup>7,9,10</sup> Brian D. Crompton,<sup>1,2,3,4,14</sup> Saud H. Aldubayan,<sup>2,5,11,12,14,\*</sup> and Eliezer M. Van Allen<sup>2,5,13,14,\*</sup>

## Summary

More knowledge is needed regarding germline predisposition to Ewing sarcoma to inform biological investigation and clinical practice. Here, we evaluated the enrichment of pathogenic germline variants in Ewing sarcoma relative to other pediatric sarcoma subtypes, as well as patterns of inheritance of these variants. We carried out European-focused and pan-ancestry case-control analyses to screen for enrichment of pathogenic germline variants in 141 established cancer predisposition genes in 1,147 individuals with pediatric sarcoma diagnoses (226 Ewing sarcoma, 438 osteosarcoma, 180 rhabdomyosarcoma, and 303 other sarcoma) relative to identically processed cancer-free control individuals. Findings in Ewing sarcoma were validated with an additional cohort of 430 individuals, and a subset of 301 Ewing sarcoma parent-proband trios was analyzed for inheritance patterns of identified pathogenic variants. A distinct pattern of pathogenic germline variants was seen in Ewing sarcoma relative to other sarcoma subtypes. *FANCC* was the only gene with an enrichment signal for heterozygous pathogenic variants in the European Ewing sarcoma discovery cohort (three individuals, OR 12.6, 95% CI 3.0–43.2,  $p = 0.003$ , FDR = 0.40). This enrichment in *FANCC* heterozygous pathogenic variants was again observed in the European Ewing sarcoma validation cohort (three individuals, OR 7.0, 95% CI 1.7–23.6,  $p = 0.014$ ), representing a broader importance of genes involved in DNA damage repair, which were also nominally enriched in individuals with Ewing sarcoma. Pathogenic variants in DNA damage repair genes were acquired through autosomal inheritance. Our study provides new insight into germline risk factors contributing to Ewing sarcoma pathogenesis.

## Introduction

Ewing sarcoma (MIM: 612219) is the second most common bone and soft tissue cancer impacting children and adolescents worldwide.<sup>1</sup> It is an aggressive malignancy that is metastatic 25% of the time at presentation and requires a very intensive treatment regimen including multiple chemotherapies as well as surgery or radiation for local control. While overall survival for localized disease has improved to 75%, treatment confers significant morbidity, and cure rates for metastatic and relapsed disease remain poor.<sup>2</sup> Through a better understanding of the predisposing genetic factors contributing to Ewing sarcoma pathogenesis, the pediatric oncology community would be able to develop more informed and less toxic treatment regimens, as well as better screen children at risk for disease, opening the door to opportunities for earlier detection and even prevention.

Ewing sarcoma is driven by *EWSR1-ETS* (MIM: 133450) gene fusions<sup>1,3</sup> and is occasionally characterized by a com-

plex rearrangement pattern known as chromoplexy.<sup>4</sup> The genetic events preceding these simple and complex rearrangements remain largely unknown. Prior work has suggested a role for pathogenic germline variants in DNA damage repair (DDR) genes in Ewing sarcoma,<sup>5,6</sup> but systematic case-control analyses to precisely define this role have not been undertaken. Much of what is known about germline predisposition to Ewing sarcoma has centered on common population variants identified as susceptibility loci from genome-wide association studies (GWASs),<sup>1,7–9</sup> and a comprehensive evaluation of the relative contribution of rare coding pathogenic germline variants is largely incomplete.

Furthermore, a more complete understanding of the familial inheritance patterns of genetic risk factors in Ewing sarcoma is needed to guide cascade testing strategies with broad potential clinical impact. While guidelines for familial testing have been developed for various cancer predisposition syndromes,<sup>10,11</sup> individuals with Ewing sarcoma and family members are not uniformly referred for

<sup>1</sup>Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA; <sup>2</sup>Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; <sup>3</sup>Department of Pediatrics, Harvard Medical School, Boston, MA 02215, USA; <sup>4</sup>Boston Children's Hospital, Boston, MA 02115, USA; <sup>5</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02115, USA; <sup>6</sup>Harvard Medical School, Boston, MA 02115, USA; <sup>7</sup>Huntsman Cancer Institute, 2000 Circle of Hope Dr, Room 4246, Salt Lake City, UT 84112, USA; <sup>8</sup>Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, 02115 MA, USA; <sup>9</sup>Department of Pediatrics, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112, USA; <sup>10</sup>Peel Therapeutics, Inc., Salt Lake City, UT 84108, USA; <sup>11</sup>Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115, USA; <sup>12</sup>College of Medicine, King Saudi Bin Abdulaziz University for Health Sciences, Riyadh 14611, Saudi Arabia; <sup>13</sup>Center for Cancer Genomics, Dana-Farber Cancer Institute, Boston, MA 02115, USA

<sup>14</sup>These authors contributed equally

\*Correspondence: [saud\\_aldubayan@dfci.harvard.edu](mailto:saud_aldubayan@dfci.harvard.edu) (S.H.A.), [eliezerm\\_vanallen@dfci.harvard.edu](mailto:eliezerm_vanallen@dfci.harvard.edu) (E.M.V.A.)

<https://doi.org/10.1016/j.ajhg.2022.04.007>

© 2022 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



genetic testing. Case reports of siblings with metachronous Ewing sarcoma diagnoses have suggested that germline variants shared within families may increase risk, but these have not yet been identified.<sup>12</sup> Family-based germline sequencing, such as the analysis of parent-proband trios, is thus a powerful tool for better understanding the inheritance of pathogenic germline variants in pediatric sarcoma generally and Ewing sarcoma in particular.<sup>13</sup>

We hypothesized that through a systematic comparative analysis of germline predisposition across pediatric sarcoma subtypes, we would elucidate distinct patterns of rare coding pathogenic variants in Ewing sarcoma. We therefore undertook a three-stage study comprising (1) European-focused and pan-ancestry case-control analyses utilizing a discovery pan-sarcoma cohort, (2) validation with an ancestry-matched case-control analysis of an additional cohort of individuals with Ewing sarcoma, and (3) evaluation of pathogenic germline variant inheritance for individuals with Ewing sarcoma utilizing parent-proband sequencing trios from a subset of the validation cohort.

## Material and methods

### Ethics approval and consent to participate

Written informed consent from individuals and institutional review board approval, allowing comprehensive genetic analysis of germline samples, were obtained by the original studies that enrolled individuals. The secondary genomic and deep-learning analyses performed for this study were approved under Dana-Farber Cancer Institute institutional review board protocols 21–143 and 20–691. This study conforms to the Declaration of Helsinki.

### Study participants

A total of 1,147 unselected individuals (879 of European ancestry) with pediatric sarcoma diagnoses were included in the discovery cohort, including 226 individuals with Ewing sarcoma (195 of European ancestry). A combination of germline whole-genome sequencing (WGS) and whole-exome sequencing (WES) was aggregated for these individuals across four data sources, and a majority of individuals were enrolled through studies associated with the St. Jude Cloud initiative.<sup>14–18</sup> We converted WGS to WES equivalents by using predefined target intervals to focus on coding variants only (supplemental methods). For validation, germline WGS for 430 individuals with Ewing sarcoma (356 of European ancestry) from the Gabriella Miller Kids First (GMKF) program was utilized, and individuals were enrolled through Project GENESIS at the Huntsman Cancer Institute and Children's Oncology Group protocol AEPI10N5. For a subset of 301 individuals with Ewing sarcoma from GMKF, germline WGS for parents was available (602 parents) and used for analysis of inheritance among trios (Tables S1 and S2; supplemental methods). Sequenced exomes for 24,128 cancer-free individuals from six cohorts were extensively quality controlled, identically processed, and analyzed in the same way as affected individuals for use as control individuals in this study (supplemental methods).

### Population stratification

We undertook principal-component analysis (PCA) with germline genotypes from sufficiently covered regions (those with 15×

coverage among 90% of samples) for all discovery, validation, and control cohorts to enable ancestry inference. We used a trained random forest classifier to assign one of the five 1000 Genomes-defined super populations (European, African, admixed American, East Asian, and South Asian) to each sample in our case and control cohorts. Affected individuals and control individuals were matched on genetic ancestry composition on the basis of the first ten principal components from the preceding analysis (supplemental methods).

### Germline variant characterization

We called germline variants with a deep learning method, DeepVariant, which has shown superior sensitivity and specificity compared with a joint genotyping-based approach (version 0.8.0).<sup>19–21</sup> High-quality coding variants were utilized for subsequent analyses (supplemental methods).

### Gene sets

We evaluated the prevalence of pathogenic variants in a list of established germline cancer predisposition genes ( $n = 141$ ; Table S3).<sup>22–25</sup> A subset of these genes had an established role in DDR ( $n = 43$ ), ascertained through evaluation of known primary biological function in the Online Mendelian Inheritance in Man (OMIM)<sup>26</sup> and Reactome<sup>27</sup> databases. The low-penetrance founder *CHEK2* (MIM: 604373) variant (p.Ile200Thr) was considered separately from other *CHEK2* pathogenic variants.

### Germline variant pathogenicity evaluation

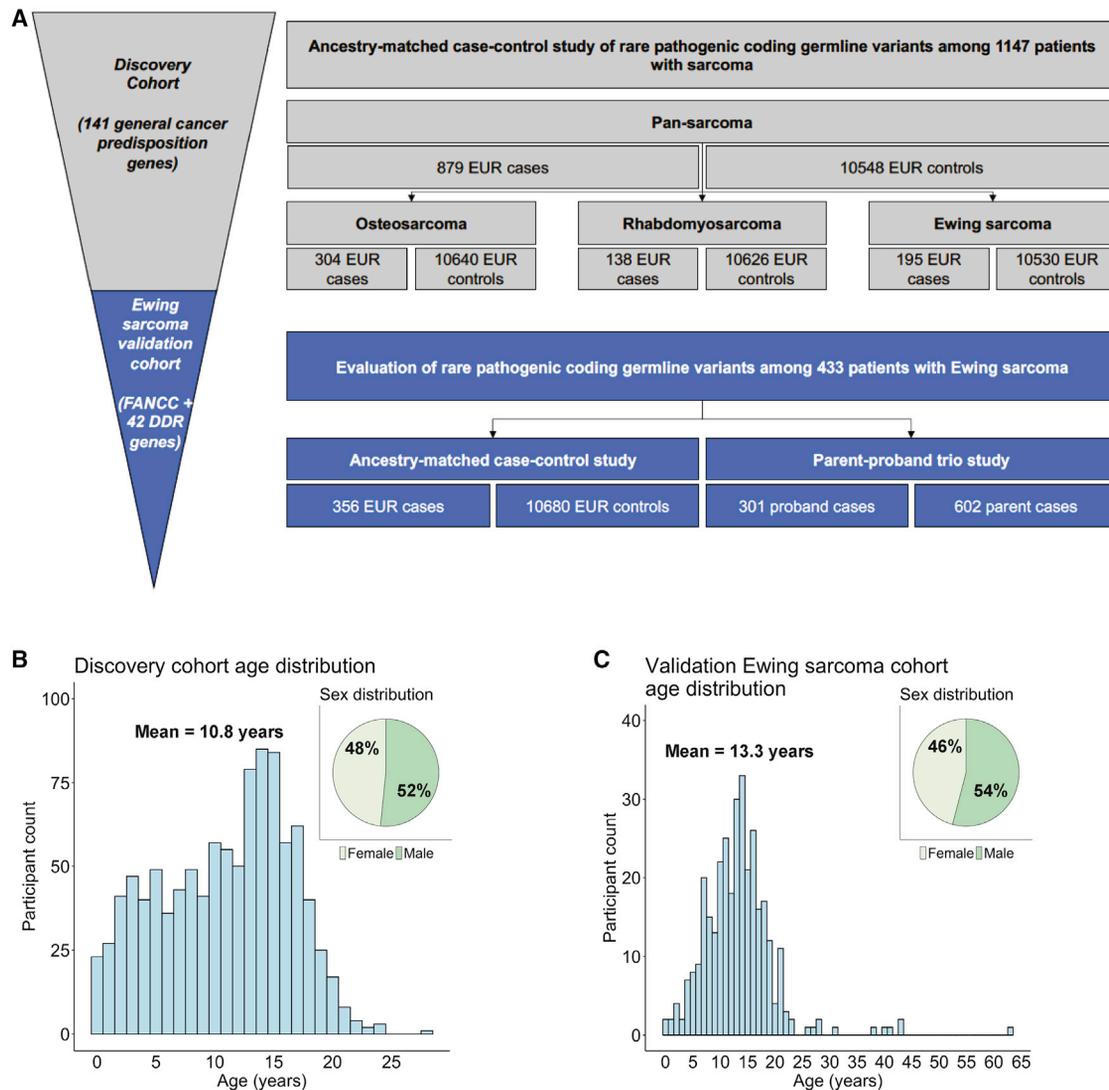
Based on ClinVar database and Variant Effect Predictor (VEP) consequence annotations, all detected germline variants in cancer predisposition genes were classified into five categories, benign, likely benign, variants of unknown significance, likely pathogenic, and pathogenic, in accordance with the American College of Medical Genetics (ACMG) guidelines.<sup>28</sup> Only putative loss-of-function, pathogenic, and likely pathogenic variants were included in this study (hereafter collectively referred to as pathogenic variants). Pathogenic variants were manually evaluated with the raw genomic data and the Integrative Genomics Viewer (IGV; supplemental methods; Tables S4, S5, and S6).<sup>29,30</sup>

### Outcomes

The primary outcomes included gene-level enrichment analysis of germline pathogenic variants in individuals with Ewing sarcoma and other pediatric sarcoma subtypes relative to cancer-free control individuals, validation of enrichment findings in Ewing sarcoma, and analysis of mechanisms of inheritance among germline pathogenic variants in Ewing sarcoma. The secondary outcomes included exploratory analysis of germline pathogenic variants in DDR genes in Ewing sarcoma.

### Statistical analysis

We used two-sided Fisher's exact tests to calculate  $p$  values. We used the R package "exact2x2" to determine the odds ratios and 95% confidence intervals (CIs) for germline pathogenic variant enrichment in affected versus cancer-free control cohorts for each of the examined cancer predisposition genes.  $p < 0.05$  was the threshold for nominal enrichment signal. For the discovery cohort, the false discovery rate (FDR) was calculated with the Benjamini-Hochberg procedure.  $FDR < 0.05$  was used as the threshold for enrichment meeting multiple hypothesis testing criteria for validation in the absence of a secondary cohort (supplemental methods).



**Figure 1. Study overview and characteristics of discovery and validation cohorts**  
 (A) Study schematic overview. The discovery cohort comprised 1,147 individuals, and the enrichment of pathogenic germline variants across 141 established cancer predisposition genes was evaluated. Case-control analyses restricted to the European ancestry (EUR) were carried out across the pan-sarcoma cohort (879 individuals), as well as major sarcoma histologic subtypes: osteosarcoma (304 individuals), rhabdomyosarcoma (138 individuals), and Ewing sarcoma (195 individuals). The validation cohort comprised 433 individuals with Ewing sarcoma. The enrichment of pathogenic germline variants in 43 DNA damage repair genes was evaluated employing an ancestry-matched case-control analysis. Mechanisms of inheritance were evaluated for 301 individuals that were a part of parent-proband trios.

(B) Discovery cohort demographics: mean age 10.8 years, 52% male.

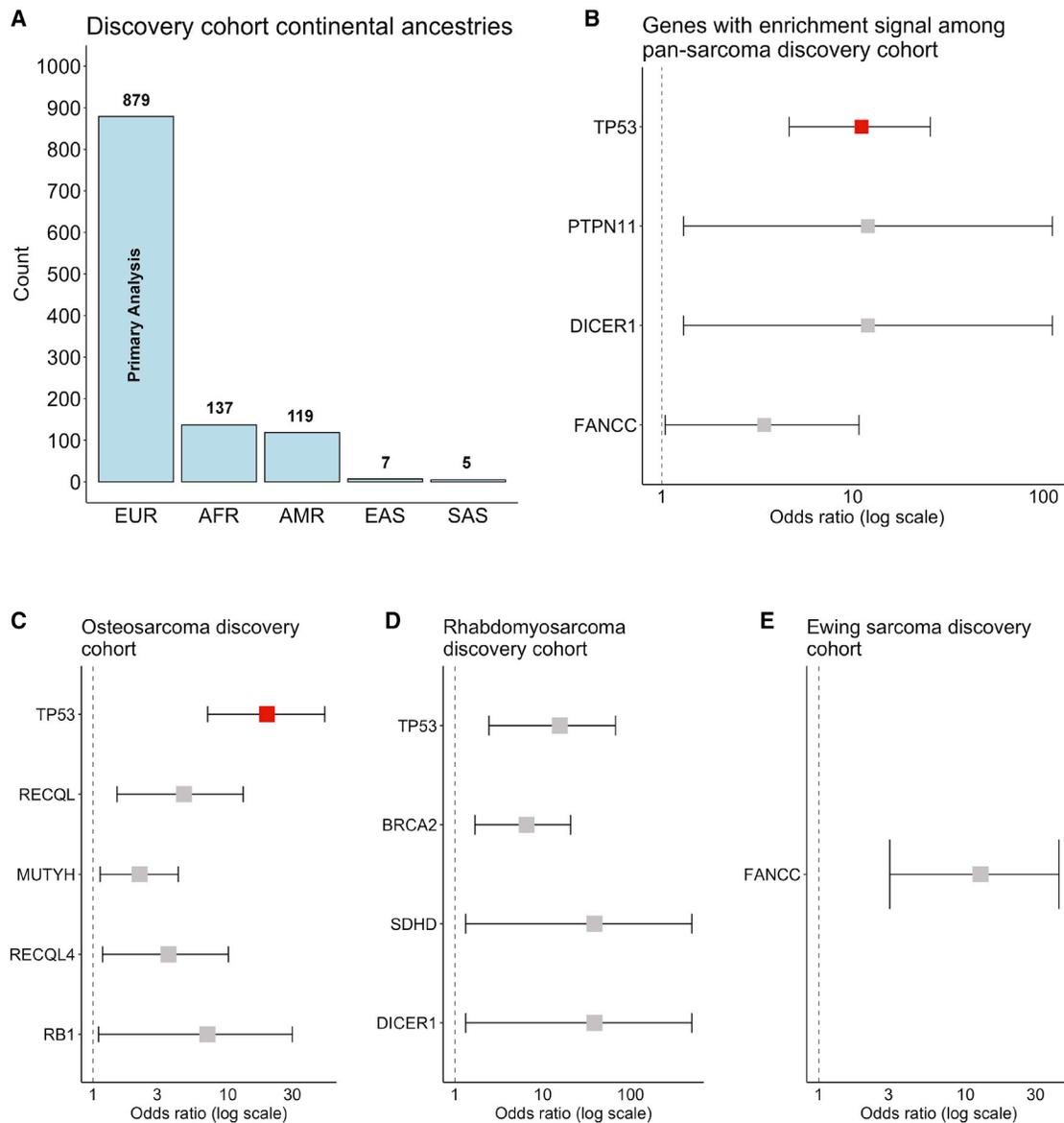
(C) Ewing sarcoma validation cohort demographics: mean age 13.3 years, 54% male.

## Results

### Study overview and characteristics of discovery and validation cohorts

Our discovery cohort of 1,147 primarily pediatric individuals with sarcoma comprised osteosarcoma (MIM: 259500) (438 individuals), rhabdomyosarcoma (MIM: 268210) (180 individuals), Ewing sarcoma (226 individuals), and other subtypes (303 individuals; [Figure 1A](#)). The mean age of individuals in the discovery cohort was 10.8 years (SD 5.5 years), and 52% of individuals were male ([Figure 1B](#)). Our validation cohort comprised 433 individuals with Ewing sarcoma. Of these, 430 individuals

were available for an ancestry-matched case-control study, and 301 individuals were available as parent-proband trios for evaluation for mechanisms of inheritance ([Figure 1A](#)). The mean age of individuals in the Ewing sarcoma validation cohort was 13.3 years (SD 6.6 years), and 54% of individuals were male ([Figure 1C](#)). The germline exome-wide mean target coverage for the discovery cohort samples was 53.9 $\times$  (interquartile range [IQR] 37.6–66.3 $\times$ ) and was 27.3 $\times$  (IQR 24.6–30.2 $\times$ ) for the Ewing sarcoma validation cohort samples. Exome-wide variant call rates were satisfactory for all samples ([Figure S1](#)). Differential coverage for evaluated genes was comparable between affected individuals and control



**Figure 2. Ancestry composition of discovery cohort**

Enrichment of pathogenic germline variants in discovery cohort, across and within major histologic subtypes

(A) Ancestry composition of discovery cohort: 879 European individuals (EUR), 137 African individuals (AFR), 119 admixed American individuals (AMR), seven East Asian individuals (EAS), and five South Asian individuals (SAS).

(B–D) Odds ratios and 95% confidence intervals for enrichment of pathogenic germline variants among cancer predisposition genes. Red: significant at  $FDR < 0.05$  after Benjamini-Hochberg procedure. Gray: significant at  $p < 0.05$  but does not meet  $FDR < 0.05$ . Genes with  $p > 0.05$  are not displayed. (B) Pan-sarcoma cohort (879 affected individuals, 10,548 control individuals). (C) Osteosarcoma subset of cohort (304 affected individuals, 10,640 control individuals). (D) Rhabdomyosarcoma subset of cohort (138 affected individuals, 10,626 control individuals).

(E) Ewing sarcoma subset of cohort (195 affected individuals, 10,530 control individuals).

individuals (Figure S2). All samples had satisfactory indel rates, variant transition-to-transversion rates, and genotype quality (Figure S3).

#### Pathogenic germline variants in cancer predisposition genes are enriched across pediatric sarcoma histologic subtypes relative to cancer-free control individuals

We assessed the frequency of pathogenic germline variants in 141 established cancer predisposition genes in our discovery cohort.<sup>22–25</sup> Our discovery cohort had representa-

tion from five major continental ancestries, and 77% of affected individuals belonged to the European ancestry. Given power considerations and the need for accurate representation of rare variant frequencies amongst matched controls, the European subset of affected individuals was utilized for the primary enrichment analysis<sup>31,32</sup> (Figure 2A; Figure S4). Control cohorts for comparison were identically processed and ancestry matched. The presence of pathogenic germline variants was not significantly associated with age (mean 10.8 versus 10.3 years,  $p = 0.29$  by

two-sided t test) or sex ( $p = 0.69$  by Fisher's exact test). Across the European pan-sarcoma discovery cohort, nominal enrichment signal at  $p < 0.05$  was observed for *TP53* (MIM: 191170) and *DICER1* (MIM: 606241), genes previously implicated in sarcoma pathogenesis.<sup>6,24,33</sup> Nominal enrichment signal was also seen for *FANCC* (MIM: 613899) and *PTPN11* (MIM: 176876), genes with less prior supporting evidence for their role in sarcoma germline predisposition. Of note, the signal in *PTPN11* was attributable to pathogenic variants seen in two individuals with soft tissue sarcoma without further histologic classification. The enrichment in *TP53* was greatest, reaching significance at  $FDR < 0.05$  across the pan-sarcoma discovery cohort (Figure 2B; Table S7).

We next carried out European ancestry-matched enrichment analyses for each of the three major sarcoma histologic subtypes within the discovery cohort. For osteosarcoma, nominal enrichment signal was observed for five genes (Table S8). These included *TP53*, *RB1* (MIM: 614041), and *RECQL4* (MIM: 603780), previously validated as important in germline predisposition to osteosarcoma (Figure 2C);<sup>34</sup> once again, only *TP53* reached significance at  $FDR < 0.05$ . The related DNA helicase gene *RECQL* (MIM: 600537) also had nominal enrichment signal in osteosarcoma-affected individuals relative to control individuals, along with *MUTYH* (MIM: 604933), genes without substantial prior evidence supporting their role in germline predisposition to osteosarcoma.<sup>35</sup>

In rhabdomyosarcoma, nominal enrichment signal was observed for *TP53* and *DICER1*, genes that have previously been implicated in germline predisposition to pediatric rhabdomyosarcoma.<sup>36</sup> We were able to redemonstrate a nominal enrichment signal for *BRCA2* (MIM: 600185), a gene with a previous moderate level of evidence for a role in germline predisposition to rhabdomyosarcoma.<sup>33,37–39</sup> We also observed a nominal enrichment signal in *SDHD* (MIM: 602690); pathogenic germline variants in succinate dehydrogenase complex genes have been reported in rhabdomyosarcoma and other sarcoma subtypes before.<sup>33</sup> No genes reached significance at  $FDR < 0.05$  (Figure 2D; Table S9).

In contrast to osteosarcoma and rhabdomyosarcoma, no pathogenic germline *TP53* variants were observed in Ewing sarcoma. Instead, in Ewing sarcoma, the only gene with nominal enrichment signal was *FANCC*, which had heterozygous pathogenic variants seen in three out of 195 individuals (1.5%, OR 12.6, 95% CI 3.0–43.2,  $p = 0.003$ ,  $FDR = 0.40$ ; Figure 2E; Figure S5; Table S10). Prior work had shown a general association between germline variants in Fanconi anemia genes and translocation-driven sarcomas, but the role of *FANCC* in germline predisposition to Ewing sarcoma had not previously been reported to our knowledge.<sup>6</sup>

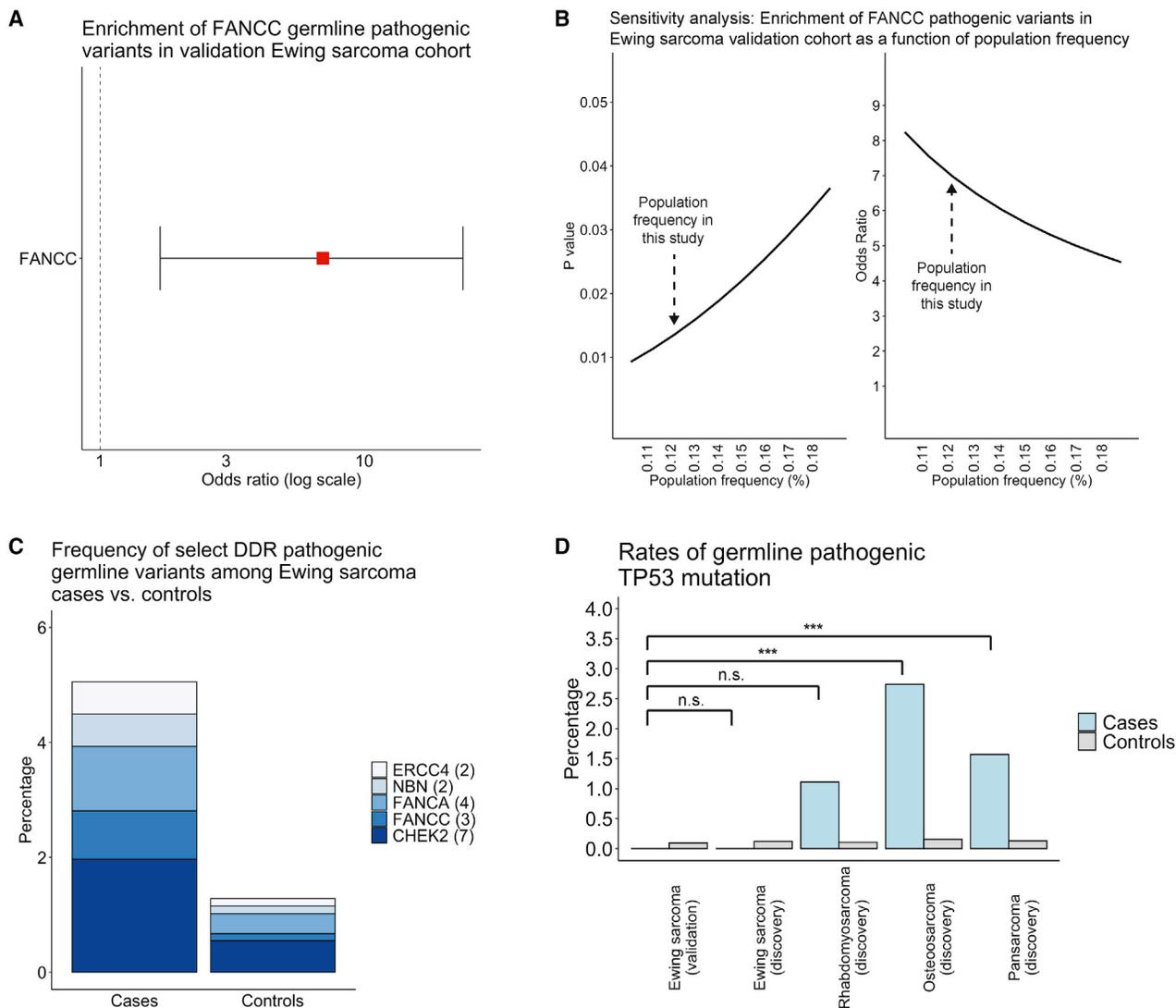
To complement the European ancestry analyses, we also carried out an additional pan-ancestry enrichment analysis including the non-European individuals from our discovery cohort (Figure S6; Tables S11, S12, S13, and S14). We

identified additional nominal enrichment signals in *NFI* (MIM: 613113), a gene with prior evidence supporting its role in sarcoma predisposition, in the pan-sarcoma and rhabdomyosarcoma cohorts.<sup>33</sup> While we were also able to recover several signals seen in our European-ancestry-focused analysis, the pan-ancestry analysis was limited by underpowering and under sampling of affected individual and control individual population frequencies. Thus, enrichment analysis of pathogenic variants in our European-ancestry discovery cohort demonstrated a unique pattern of predisposing variants across pediatric sarcoma subtypes, with a strong enrichment signal for *TP53* in all sarcoma subtypes except Ewing sarcoma.

#### ***FANCC* and other DNA damage repair genes harbor pathogenic germline variants in Ewing sarcoma**

Having demonstrated a distinct enrichment pattern among individuals with Ewing sarcoma relative to those with osteosarcoma or rhabdomyosarcoma, we proceeded to evaluate the enrichment signal in *FANCC* in our larger validation cohort of Ewing sarcoma-affected individuals. European individuals from our independent Ewing sarcoma validation cohort were also ancestry matched to cancer-free control individuals to enable targeted evaluation of *FANCC* enrichment (Figure S7). Heterozygous pathogenic germline *FANCC* variants were again enriched, seen in three of 356 individuals (0.8%, OR 7.0, 95% CI 1.7–23.6,  $p = 0.014$ , single hypothesis; Figure 3A). The population frequency of pathogenic germline *FANCC* variants amongst our matched European controls was 0.12%. Population frequencies of pathogenic germline *FANCC* variants in the non-Finnish European ancestry have been reported in the range of 0.10%–0.18% in the literature.<sup>24,40,41</sup> We thus carried out a sensitivity analysis by varying the number of matched controls with pathogenic *FANCC* variants over this range of frequencies and found that the enrichment of pathogenic germline *FANCC* variants in the Ewing sarcoma validation cohort would remain significant (Figure 3B). The pooled odds ratio for *FANCC* enrichment between discovery and validation cohorts was also significant (six of 551 individuals, 1.1%, OR 9.0, 95% CI 3.7–22.0,  $p < .0001$ ).

Prior mechanistic work has demonstrated that *FANCC* knockout contributes to rearrangement signatures consistent with homologous recombination deficiency.<sup>42</sup> We thus asked whether the recurrent enrichment of *FANCC* represented a broader importance of DDR genes in germline predisposition to Ewing sarcoma, as has been previously suggested.<sup>5</sup> We performed an exploratory analysis in the Ewing sarcoma validation cohort on the subset of 43 cancer predisposition genes with specific DDR roles. We identified seven pathogenic germline *CHEK2* variants (OR 3.6, 95% CI 1.6–7.9,  $p = 0.005$ ) and four pathogenic germline *FANCA* (MIM: 607139) variants (OR 3.3, 95% CI 1.1–9.1,  $p = 0.042$ ) contributing to nominal enrichment in these genes in affected individuals relative to control individuals. Marginal signals in *ERCC4* (MIM: 133520) (two individuals, OR 4.3, 95% CI 0.7–18,  $p = 0.09$ ) and



**Figure 3. Enrichment of pathogenic germline variants in *FANCC* and other DNA damage repair genes in Ewing sarcoma validation cohort**

(A) Enrichment of pathogenic germline variants in *FANCC* in the Ewing Sarcoma validation cohort versus control individuals (OR 7.0, 95% CI 1.7–23.6,  $p = 0.014$ ).

(B) Sensitivity analysis for enrichment of pathogenic germline *FANCC* variants in the Ewing sarcoma validation cohort over a range of simulated population frequencies.

(C) Collective frequency of pathogenic variants in leading gene set of DNA damage repair genes in the Ewing sarcoma validation cohort. In addition to *FANCC*, nominal enrichment signals were also seen in *CHEK2* (OR 3.6, 95% CI 1.6–7.9,  $p = 0.005$ ) and *FANCA* (OR 3.3, 95% CI 1.1–9.1,  $p = 0.042$ ).

(D) Rates of pathogenic variants in *TP53* in Ewing sarcoma validation cohort in comparison to Ewing sarcoma subset of discovery cohort, rhabdomyosarcoma subset of discovery cohort, osteosarcoma subset of discovery cohort, and pan-sarcoma discovery cohort (Fisher's exact tests, n.s. denotes no significant difference, \*\*\* denotes significant difference at  $p < 0.05$ ).

*NBN* (MIM: 602667) (two individuals, OR 4.3, 95% CI 0.7–18,  $p = 0.09$ ) were also seen. In combination with *FANCC*, these five genes harbored germline pathogenic variants in 18 of 356 Ewing sarcoma-affected individuals (5.1%) compared with 137 of 10,680 control individuals (1.3%; Figure 3C; Table S15).

The signals in *FANCC*, *FANCA*, and *CHEK2* were seen in the European ancestry subset of the Ewing sarcoma validation cohort, and our pan-ancestry analysis was thus underpowered to recover many of these signals (Figure S8; Table S16). Across all individuals with Ewing sarcoma in

the validation cohort, the presence of pathogenic germline variants in DDR genes was not significantly associated with age (mean 13.3 versus 13.6 years,  $p = 0.75$  by two-sided  $t$  test) or sex ( $p = 0.15$  by Fisher's exact test).

Similar to our discovery Ewing sarcoma cohort, we once again identified no pathogenic germline *TP53* variants in our validation Ewing sarcoma cohort. The rate of pathogenic germline *TP53* variants in individuals with Ewing sarcoma (0%) was significantly lower than that seen for all sarcomas in aggregate (1.6%,  $p = 0.006$  by Fisher's exact test) and osteosarcoma in particular (2.7%,  $p = 0.0005$  by

Fisher's exact test; Figure 3D). This, in combination with the recurrent enrichment of pathogenic germline variants in *FANCC*, as well as the nominal enrichment of pathogenic germline variants in *CHEK2* and *FANCA*, demonstrated a distinct pattern of germline variants in Ewing sarcoma relative to other pediatric sarcoma subtypes.

### Pathogenic germline variants in DNA damage repair genes are inherited in high-risk families

Having identified pathogenic germline variants in *FANCC* and other DDR genes in Ewing sarcoma, we next sought to assess inheritance of these variants. Thus, we evaluated the 301 individuals with Ewing sarcoma from our validation cohort that were part of parent-proband sequencing trios. Among these 301 individuals, 32 harbored pathogenic germline variants in DDR genes (10.6%; Figure 4A). In 32 of 32 probands in which a pathogenic germline DDR variant was identified in a proband, the same germline DDR variant was identified in one of the parents (100%). In contrast, for probands in whom a pathogenic germline DDR variant was not identified, only 19 of 269 had at least one parent with a germline DDR variant (7.1%; Figure 4B). While pathogenic germline variants in genes such as *BRCA2* and *CHEK2* were also observed in some parents and not inherited by probands, these were at a rate that was comparable to the population frequency (Figure S9).

Identical pathogenic germline DDR variants in probands and parents impacted *FANCC*, *ERCC2* (MIM: 126340), *CHEK2*, and *BRCA1* (MIM: 113705) among other genes (Figure 4C). In three instances, heterozygous germline pathogenic variants affecting multiple DDR genes were seen in probands and each variant was also identified in a parent (Figure 4D).

We sought to understand whether as yet unidentified *de novo* pathogenic variants in other coding genes may coordinate with or complement the inherited DDR variants to explain a significant proportion of the unexplained germline risk for developing pediatric Ewing sarcoma. Based on prior frameworks,<sup>43–45</sup> we reasoned that finding pathogenic *de novo* variants recurrently impacting the same gene in a cohort of 301 proband-parent trios would be highly unlikely by chance, implicating potential additional candidate risk genes. However, in our cohort, we identified recurrent pathogenic *de novo* germline variants in only one gene, *TTN* (MIM: 188840), which occurred in two separate individuals with Ewing sarcoma. As the frequency of pathogenic germline variants in *TTN* between affected individuals and cancer-free control individuals was not significantly different and there is no established biological role for *TTN* in Ewing sarcoma oncogenesis, there was insufficient evidence to support its role in germline predisposition to Ewing sarcoma (Figure S10).

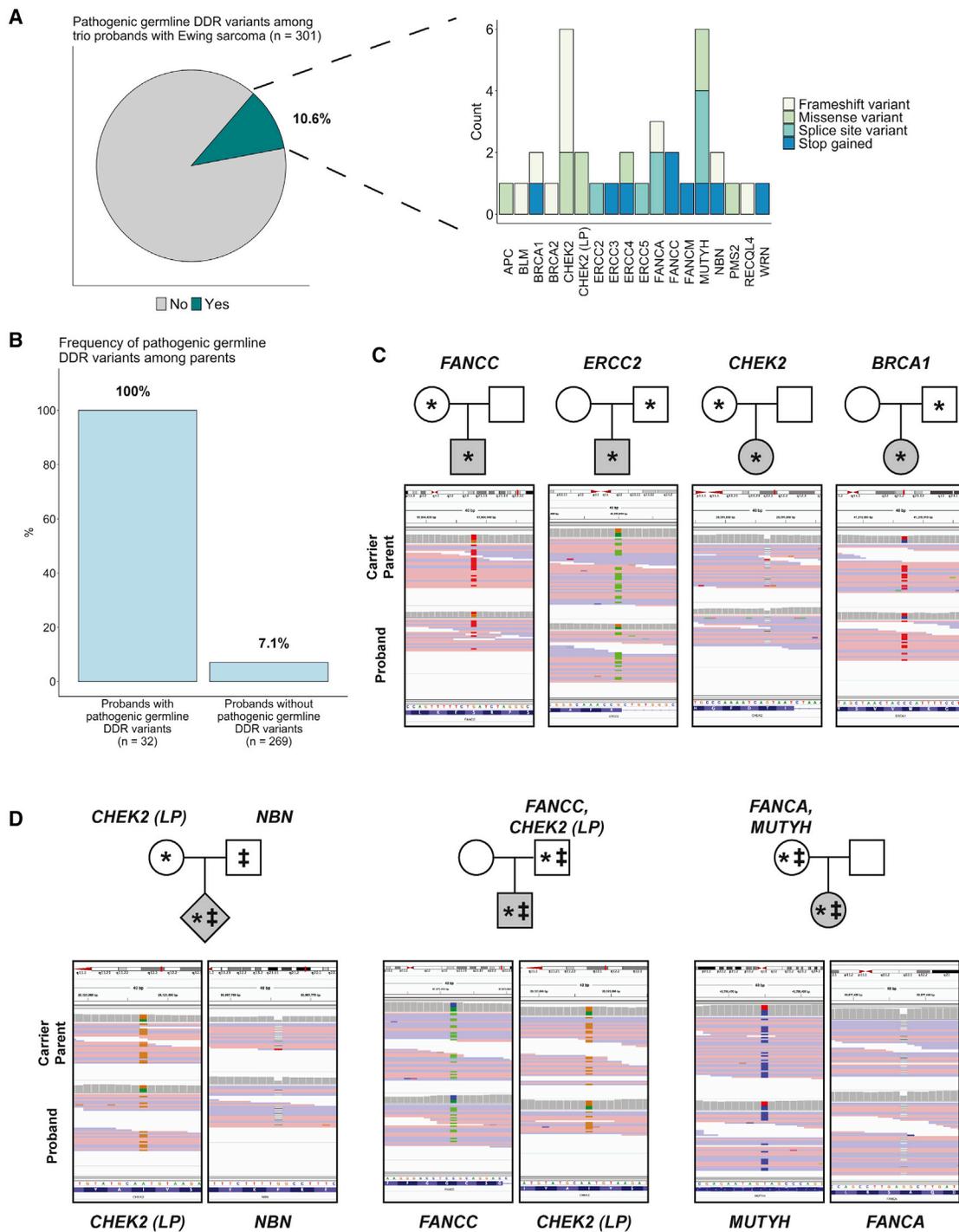
Taken together, pathogenic germline variants in DDR genes were frequently observed in families of individuals with Ewing sarcoma. Autosomal inheritance, as opposed to *de novo* development, was the mechanism of inheritance of these moderate penetrance risk variants.

## Discussion

This study represents a systematic analysis of germline predisposition to Ewing sarcoma relative to other pediatric sarcoma subtypes. Having assembled germline sequencing data from 1,147 individuals with pediatric sarcoma diagnoses, we undertook European-focused and pan-ancestry case-control analyses and illustrated distinct patterns of enrichment amongst pathogenic variants in Ewing sarcoma relative to osteosarcoma and rhabdomyosarcoma. Supporting the validity of our approach, we were able to recover enrichment signal in many cancer predisposition genes known to be associated with pediatric sarcoma risk in the European-focused analysis, such as *TP53*, *RB1*, and *DICER1*. We additionally demonstrated enrichment signal in cancer predisposition genes with less well-characterized links to pediatric sarcoma, most notably *FANCC* in Ewing sarcoma. Our pan-ancestry analysis was able to identify a shared enrichment signal in pathogenic germline variants in *NF1* across ancestries, although this approach was otherwise limited by its underpowering and under sampling of baseline population frequencies.

We then validated the enrichment of pathogenic germline variants in *FANCC* among individuals with Ewing sarcoma by using an independent cohort. This recurrent enrichment of heterozygous pathogenic germline variants in *FANCC* provides evidence for its role in increasing risk for some individuals with Ewing sarcoma and raises the possibility that monoallelic germline variants in Fanconi anemia genes may confer increased risk in other translocation-associated cancers. We demonstrated that the enrichment in *FANCC* pathogenic germline variants represented a broader importance of DDR genes, including *FANCA* and *CHEK2*. While prior studies have identified occasional instances of pathogenic germline *TP53* variants amongst individuals with Ewing sarcoma, through comparative analyses, we found that the frequency of pathogenic germline *TP53* variants among individuals with Ewing sarcoma was significantly lower in relation to other pediatric sarcoma subtypes.<sup>5,37</sup> This finding is supported by the clinical observation that Ewing sarcoma is not frequently seen in families with Li-Fraumeni syndrome.<sup>46</sup>

Using parent-proband trios, we showed that pathogenic germline variants in DNA damage repair genes found in individuals with Ewing sarcoma are also present in their parents and therefore passed on through autosomal inheritance. As moderate penetrance risk variants that are also present in parents, we reasoned that pathogenic germline variants in some DDR genes play a substantial role in increasing risk for developing Ewing sarcoma but are most likely not sufficient to cause the disease in isolation. However, we did not identify *de novo* variants recurrently impacting other genes to support a role for their interaction with pathogenic germline variants in DDR genes to promote germline predisposition to Ewing sarcoma.



**Figure 4. Pathogenic germline variants in DNA damage repair genes are inherited in high-risk families**

(A) Pathogenic germline variants in DDR genes among trio probands with Ewing sarcoma. 35 pathogenic variants impacting 32 of 301 individuals with Ewing sarcoma were identified.

(B) 32 of 32 probands (100%) with pathogenic germline variants in DDR genes had identical variants identified in parents. 19 of 269 probands without a pathogenic germline variant in a DDR gene had at least one parent with a germline DDR variant that was not inherited by the proband (7.1%).

(C and D) Pedigrees and IGV screenshots of pathogenic variants in DDR genes.

Pedigree legend: circle, female sex; square, male sex; diamond, unknown sex; gray shading, proband with Ewing sarcoma; \* denotes variant 1 identified in parent-proband trio; ‡ denotes variant 2 identified in parent-proband trio. Top screenshot: carrier parent. Bottom screenshot: proband.

Our study had some notable limitations. First, we were limited in recovering enrichment signals in non-European individuals because of the underpowering of case and control cohorts in other ancestries. Second, while pathogenic germline variants in *FANCC* occurred at a rate greater than expected by chance among individuals with Ewing sarcoma, the overall frequency of these variants was low (1.5% in the European discovery cohort, 0.8% in the European validation cohort), supporting the role of *FANCC* as a moderate penetrance cancer predisposition gene as opposed to the sole driver of disease pathogenesis. Based on the recurrent signals in *FANCC*, and modest signals in other DNA damage repair genes in the validation cohort in the absence of pathogenic variants in highly penetrant genes such as *TP53*, these variants are likely insufficient to lead to Ewing sarcoma in isolation and potentially coordinate with other germline and oncogenic processes to incrementally increase risk. Additionally, our focus on germline variants in select DDR genes within Ewing sarcoma most likely underestimated the total contribution of rare coding pathogenic germline variants to Ewing sarcoma pathogenesis. Finally, similar to much preceding work in germline predisposition in pediatric cancers, our methods were limited to identifying known pathogenic germline SNVs/indels conferring increased risk in pediatric sarcoma. As progress is made in germline structural variant discovery<sup>47</sup> and placing rare pathogenic variants in the context of complex germline interactions,<sup>48–51</sup> future studies and new statistical frameworks will be needed to more completely define the role of germline predisposition in Ewing sarcoma pathogenesis.

Taken together, our analysis supports a unique contribution of germline variants in *FANCC* and other DDR genes to Ewing sarcoma pathogenesis. Our study provides a foundation to inform approaches to genetic testing for individuals with Ewing sarcoma as well as cascade testing for family members. Future studies on independent and ancestrally diverse cohorts should be undertaken to further evaluate the link between pathogenic germline variants in DDR genes and Ewing sarcoma.

### Data and code availability

Sequencing data analyzed in this study were obtained from multiple publicly accessible data repositories. Sequenced samples from the discovery cohort came from four data sources: St. Jude Cloud (Pediatric Cancer Genome Project, St. Jude Lifetime, Genomes for Kids, and Childhood Cancer Survivor Study;  $n = 1,033$ ),<sup>14–18</sup> dbGaP study “Genomic Sequencing of Ewing Sarcoma” (dbGaP: phs000804.v1.p1;  $n = 26$ ),<sup>52,53</sup> dbGaP study “Osteosarcoma Genomics” (dbGaP: phs000699.v1.p1,  $n = 58$ ),<sup>54</sup> and ICGC study “Bone Cancer – UK” (BOCA-UK,  $n = 63$ ).<sup>55</sup> Sequenced samples from the Ewing sarcoma validation and trio cohorts were from the Gabriella Miller Kids First “Ewing Sarcoma – Genetic Risk” study (dbGaP: phs001228.v1.p1).<sup>56</sup> The control cohorts came from the following sources: Autism Sequencing Consortium (dbGaP: phs000298), Framingham Cohort (dbGaP: phs000007), Multi-Ethnic Study of Atherosclerosis (dbGaP: phs000209), Lung Cohort (dbGaP: phs000291), in-house collection of exomes from National Heart,

Lung, and Blood Institute “Grand Opportunity” Exome Sequencing Project (NHLBI GO-ESP), and the 1000 Genomes Project.<sup>57</sup>

### Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.ajhg.2022.04.007>.

### Acknowledgments

This work was supported by a research training grant in Pediatric Oncology NIH 5T32CA136432-12 (R.G.), American Society of Clinical Oncology Conquer Cancer – Sarcoma Foundation of America young investigator award (R.G.), Jay Vernon Jackson Memorial Research Award from Sarcoma Foundation of America (R.G.), T32GM007753 from the National Institute of General Medical Sciences (J.K.J.), NIH R01 CA227237 (A.G.), NIH R21 HG010748 (A.G.), Edward B. Clark, MD Endowed Chair in Pediatric Research (J.D.S.), Helen Clise Presidential Endowed Chair in Li-Fraumeni Syndrome Research (J.D.S.), Conquer Cancer Foundation Career Development Award 13167 from the American Society of Clinical Oncology (S.H.A.), young investigator award 18YOUN02 from the Prostate Cancer Foundation (S.H.A.), Physician Research Award (PC200150/W81XWH-21-1-0084) of the Department of Defense (S.H.A.), NIH R37 CA222574 (E.M.V.A.), NIH R01 CA227388 (E.M.V.A.), NIH U01 CA233100 (E.M.V.A.), Innovation in Cancer Informatics Award (E.M.V.A.), COG Biospecimen Bank grant U24CA196173, NCTN Operations Center grant U10CA180886, NCTN Statistics and Data Center grant U10CA180899, and Gabriella Miller Kids First 1X01 HL 140547-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

### Author contributions

R.G., E.M.V.A., and S.H.A. conceived the project design, carried out the analysis, and wrote the manuscript. B.D.C. and K.A.J. contributed to refinement of the project design, data interpretation, validation of findings, and manuscript review. S.Y.C., S.H., J.K.J., and H.C. provided support with data analysis, data interpretation, and manuscript review. J.D.S. contributed to validation of findings, data interpretation, and manuscript review. E.Y., S.O., L.H., G.M., and T.F. carried out foundational work that enabled validation of findings. J.K. and A.G. contributed to data interpretation and manuscript review.

### Declaration of interests

R.G. has equity in Google, Microsoft, Amazon, Apple, Moderna, Pfizer, and Vertex Pharmaceuticals. J.K.’s spouse has received consulting fees from ROME Therapeutics, Foundation Medicine, Inc., NanoString Technologies, EMD Millipore Sigma, Pfizer, and Tekla Capital; is a founder and has equity in ROME Therapeutics, PanTher Therapeutics, and TellBio, Inc.; and receives research support from ACD-Biotechnie, PureTech Health LLC, and Ribon Therapeutics. J.D.S. is co-founder and shareholder in ItRunsInMyFamily.com and is co-founder, shareholder, and employed by Peel Therapeutics, Inc. E.M.V.A. holds consulting roles with Tango Therapeutics, Genome Medical, Genomic Life, Enara Bio, Janssen, and Manifold

Bio; he receives research support from Bristol-Myers Squibb and Novartis; he has equity in Tango Therapeutics, Genome Medical, Genomic Life, Syapse, Enara Bio, Manifold Bio, and Microsoft; he has received travel reimbursement from Roche and Genentech; and he has filed institutional patents on chromatin mutations, immunotherapy response, and methods for clinical interpretation. The other authors declare no competing interests.

Received: January 4, 2022

Accepted: April 11, 2022

Published: May 4, 2022

## Web resources

OMIM, <http://www.omim.org/>

## References

1. Grünewald, T.G.P., Cidre-Aranaz, F., Surdez, D., Tomazou, E.M., De Álava, E., Kovar, H., Sorensen, P.H., Delattre, O., and Dirksen, U. (2018). Ewing sarcoma. *Nat. Rev. Dis. Prim.* **4**, 5. <https://doi.org/10.1038/s41572-018-0003-x>.
2. Balamuth, N.J., and Womer, R.B. (2010). Ewing's sarcoma. *Lancet Oncol.* **11**, 184–192. [https://doi.org/10.1016/s1470-2045\(09\)70286-4](https://doi.org/10.1016/s1470-2045(09)70286-4).
3. Brohl, A.S., Solomon, D.A., Chang, W., Wang, J., Song, Y., Sindiri, S., Patidar, R., Hurd, L., Chen, L., Shern, J.F., et al. (2014). The genomic landscape of the ewing sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet.* **10**, e1004475. <https://doi.org/10.1371/journal.pgen.1004475>.
4. Anderson, N.D., De Borja, R., Young, M.D., Fuligni, F., Rosic, A., Roberts, N.D., Hajjar, S., Layeghifard, M., Novokmet, A., Kowalski, P.E., et al. (2018). Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors. *Science* **361**, 361. <https://doi.org/10.1126/science.aam8419>.
5. Brohl, A.S., Patidar, R., Turner, C.E., Wen, X., Song, Y.K., Wei, J.S., Calzone, K.A., and Khan, J. (2017). Frequent inactivating germline mutations in DNA repair genes in patients with Ewing sarcoma. *Genet. Med.* **19**, 955–958. <https://doi.org/10.1038/gim.2016.206>.
6. Ballinger, M.L., Goode, D.L., Ray-Coquard, I., James, P.A., Mitchell, G., Niedermayr, E., Puri, A., Schiffman, J.D., Dite, G.S., Cipponi, A., et al. (2016). Monogenic and polygenic determinants of sarcoma risk: an international genetic study. *Lancet Oncol.* **17**, 1261–1271. [https://doi.org/10.1016/s1470-2045\(16\)30147-4](https://doi.org/10.1016/s1470-2045(16)30147-4).
7. Grünewald, T.G.P., Bernard, V., Gilardi-Hebenstreit, P., Raynal, V., Surdez, D., Aynaud, M.M., Mirabeau, O., Cidre-Aranaz, F., Tirode, F., Zaidi, S., et al. (2015). Chimeric EWSR1-FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite. *Nat. Genet.* **47**, 1073–1078. <https://doi.org/10.1038/ng.3363>.
8. Postel-Vinay, S., Véron, A.S., Tirode, F., Pierron, G., Reynaud, S., Kovar, H., Oberlin, O., Lapouble, E., Ballet, S., Lucchesi, C., et al. (2012). Common variants near TARDBP and EGR2 are associated with susceptibility to Ewing sarcoma. *Nat. Genet.* **44**, 323–327. <https://doi.org/10.1038/ng.1085>.
9. Machiela, M.J., Grünewald, T.G.P., Surdez, D., Reynaud, S., Mirabeau, O., Karlins, E., Rubio, R.A., Zaidi, S., Grossetete-Lalami, S., Ballet, S., et al. (2018). Genome-wide association study identifies multiple new loci associated with Ewing sarcoma susceptibility. *Nat. Commun.* **9**, 3184–3188. <https://doi.org/10.1038/s41467-018-05537-2>.
10. Kratz, C.P., Achatz, M.I., Brugieres, L., Frebourg, T., Garber, J.E., Greer, M.L.C., Hansford, J.R., Janeway, K.A., Kohlmann, W.K., McGee, R., et al. (2017). Cancer screening recommendations for individuals with Li-Fraumeni syndrome. *Clin. Cancer Res.* **23**, e38–e45. <https://doi.org/10.1158/1078-0432.ccr-17-0408>.
11. Schultz, K.A.P., Williams, G.M., Kamihara, J., Stewart, D.R., Harris, A.K., Bauer, A.J., Turner, J., Shah, R., Schneider, K., Schneider, K.W., et al. (2018). Dicer1 and associated conditions: identification of at-risk individuals and recommended surveillance strategies. *Clin. Cancer Res.* **24**, 2251–2261. <https://doi.org/10.1158/1078-0432.ccr-17-3089>.
12. Joyce, M.J., Harmon, D.C., Mankin, H.J., Suit, H.D., Schiller, A.L., and Truman, J.T. (1984). Ewing's sarcoma in female siblings: a clinical report and review of the literature. *Cancer* **53**, 1959–1962. [https://doi.org/10.1002/1097-0142\(19840501\)53:9<1959::aid-cnrc2820530926>3.0.co;2-9](https://doi.org/10.1002/1097-0142(19840501)53:9<1959::aid-cnrc2820530926>3.0.co;2-9).
13. Kuhlén, M., Taeubner, J., Brozou, T., Wiczorek, D., Siebert, R., and Borkhardt, A. (2019). Family-based germline sequencing in children with cancer. *Oncogene* **38**, 1367–1380. <https://doi.org/10.1038/s41388-018-0520-9>.
14. McLeod, C., Gout, A.M., Zhou, X., Thrasher, A., Rahbarinia, D., Brady, S.W., Macias, M., Birch, K., Finkelstein, D., Sunny, J., et al. (2021). St. Jude cloud: a pediatric cancer genomic data-sharing ecosystem. *Cancer Discov.* **11**, 1082–1099. <https://doi.org/10.1158/2159-8290.cd-20-1230>.
15. Chen, X., Stewart, E., Shelat, A.A., Qu, C., Bahrami, A., Hatley, M., Wu, G., Bradley, C., McEvoy, J., Pappo, A., et al. (2013). Targeting oxidative stress in embryonal rhabdomyosarcoma. *Cancer Cell* **24**, 710–724. <https://doi.org/10.1016/j.ccr.2013.11.002>.
16. Downing, J.R., Wilson, R.K., Zhang, J., Mardis, E.R., Pui, C.H., Ding, L., Ley, T.J., and Evans, W.E. (2012). The pediatric cancer genome project. *Nat. Genet.* **44**, 619–622. <https://doi.org/10.1038/ng.2287>.
17. Wang, Z., Wilson, C.L., Easton, J., Thrasher, A., Mulder, H., Liu, Q., Hedges, D.J., Wang, S., Rusch, M.C., Edmonson, M.N., et al. (2018). Genetic risk for subsequent neoplasms among long-term survivors of childhood cancer. *J. Clin. Oncol.* **36**, 2078–2087. <https://doi.org/10.1200/jco.2018.77.8589>.
18. Robison, L.L., Mertens, A.C., Boice, J.D., Breslow, N.E., Donaldson, S.S., Green, D.M., Li, F.P., Meadows, A.T., Mulvihill, J.J., Neglia, J.P., et al. (2002). Study design and cohort characteristics of the Childhood Cancer Survivor Study: a multi-institutional collaborative project. *Med. Pediatr. Oncol.* **38**, 229–239. <https://doi.org/10.1002/mpo.1316>.
19. Poplin, R., Chang, P.C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., et al. (2018). A universal snp and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **36**, 983–987. <https://doi.org/10.1038/nbt.4235>.
20. AlDubayan, S.H., Conway, J.R., Camp, S.Y., Witkowski, L., Kofman, E., Reardon, B., Han, S., Moore, N., Elmarakeby, H., Salari, K., et al. (2020). Detection of pathogenic variants with germline genetic testing using deep learning vs standard methods in patients with prostate cancer and melanoma. *JAMA* **324**, 1957–1969. <https://doi.org/10.1001/jama.2020.20457>.
21. Camp, S.Y., Kofman, E., Reardon, B., Moore, N.D., Al-Rubaish, A.M., Aljumaan, M., Al-Ali, A.K., Van Allen, E.M., Taylor-Weiner, A., and AlDubayan, S.H. (2021). Evaluating the

- molecular diagnostic yield of joint genotyping-based approach for detecting rare germline pathogenic and putative loss-of-function variants. *Genet. Med.* 23, 918–926. <https://doi.org/10.1038/s41436-020-01074-w>.
22. Rahman, N. (2014). Realizing the promise of cancer predisposition genes. *Nature* 505, 302–308. <https://doi.org/10.1038/nature12981>.
  23. Huang, K.L., Mashl, R.J., Wu, Y., Ritter, D.I., Wang, J., Oh, C., Paczkowska, M., Reynolds, S., Wyczalkowski, M.A., Oak, N., et al. (2018). Pathogenic germline variants in 10,389 adult cancers. *Cell* 173, 355–370. <https://doi.org/10.1016/j.cell.2018.03.039>.
  24. Mirabello, L., Zhu, B., Koster, R., Karlins, E., Dean, M., Yeager, M., Gianferante, M., Spector, L.G., Morton, L.M., Karyadi, D., et al. (2020). Frequency of pathogenic germline variants in cancer-susceptibility genes in patients with osteosarcoma. *JAMA Oncol.* 6, 724–734. <https://doi.org/10.1001/jamaoncol.2020.0197>.
  25. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47, D941–D947. <https://doi.org/10.1093/nar/gky1015>.
  26. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 43, D789–D798. <https://doi.org/10.1093/nar/gku1205>.
  27. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503. <https://doi.org/10.1093/nar/gkz1031>.
  28. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., 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. <https://doi.org/10.1038/gim.2015.30>.
  29. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. <https://doi.org/10.1038/nbt.1754>.
  30. Robinson, J.T., Thorvaldsdóttir, H., Wenger, A.M., Zehir, A., and Mesirov, J.P. (2017). Variant review with the integrative genomics viewer. *Cancer Res.* 77, e31–e34. <https://doi.org/10.1158/0008-5472.can-17-0337>.
  31. Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., and Bustamante, C.D. (2011). Demographic history and rare allele sharing among human populations. *Proc. Natl. Acad. Sci. U S A* 108, 11983–11988. <https://doi.org/10.1073/pnas.1019276108>.
  32. Petrovski, S., and Goldstein, D.B. (2016). Unequal representation of genetic variation across ancestry groups creates health-care inequality in the application of precision medicine. *Genome Biol.* 17, 157–218. <https://doi.org/10.1186/s13059-016-1016-y>.
  33. Li, H., Sisouduya, S.D., Martin-Giacalone, B.A., Khayat, M.M., Dugan-Perez, S., Marquez-Do, D.A., Scheurer, M.E., Muzny, D., Boerwinkle, E., Gibbs, R.A., et al. (2020). Germline cancer predisposition variants in pediatric rhabdomyosarcoma: a report from the Children’s oncology Group. *JNCI J. Natl. Cancer Inst.* 00, 1–9. <https://doi.org/10.1093/jnci/djaa204>.
  34. Gianferante, D.M., Mirabello, L., and Savage, S.A. (2017). Germline and somatic genetics of osteosarcoma - connecting aetiology, biology and therapy. *Nat. Rev. Endocrinol.* 13, 480–491. <https://doi.org/10.1038/nrendo.2017.16>.
  35. Mirabello, L., Yu, K., Berndt, S.I., Burdett, L., Wang, Z., Chowdhury, S., Teshome, K., Uzoka, A., Hutchinson, A., Grotmol, T., et al. (2011). A comprehensive candidate gene approach identifies genetic variation associated with osteosarcoma. *BMC Cancer* 11, 209. <https://doi.org/10.1186/1471-2407-11-209>.
  36. Martin-Giacalone, B.A., Weinstein, P.A., Plon, S.E., and Lupo, P.J. (2021). Pediatric rhabdomyosarcoma: epidemiology and genetic susceptibility. *J. Clin. Med.* 10, 2028. <https://doi.org/10.3390/jcm10092028>.
  37. Zhang, J., Walsh, M.F., Wu, G., Edmonson, M.N., Gruber, T.A., Easton, J., Hedges, D., Ma, X., Zhou, X., Yergeau, D.A., et al. (2015). Germline mutations in predisposition genes in pediatric cancer. *N. Engl. J. Med.* 373, 2336–2346. <https://doi.org/10.1056/nejmoa1508054>.
  38. Kim, J., Light, N., Subasri, V., Young, E.L., Wegman-Ostrosky, T., Barkauskas, D.A., Hall, D., Lupo, P.J., Patidar, R., Maese, L.D., et al. (2021). Pathogenic germline variants in cancer susceptibility genes in children and Young adults with rhabdomyosarcoma. *JCO Precis. Oncol.*, 75–87. <https://doi.org/10.1200/po.20.00218>.
  39. Naslund-Koch, C., Nordestgaard, B.G., and Bojesen, S.E. (2016). Increased risk for other cancers in addition to breast cancer for CHEK2\*1100delC heterozygotes estimated from the copenhagen general population study. *J. Clin. Oncol.* 34, 1208–1216. <https://doi.org/10.1200/jco.2015.63.3594>.
  40. Article, O. (2021). Breast cancer risk genes — Association analysis in more than 113,000 women. *N. Engl. J. Med.* 384, 428–439. <https://doi.org/10.1056/NEJMoa1913948>.
  41. Karczewski, K.J., Francioli, L.C., and MacArthur, D.G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Yearb. Paediatr. Endocrinol.* 581, 434–443. <https://doi.org/10.1530/ey.17.14.3>.
  42. Zou, X., Owusu, M., Harris, R., Jackson, S.P., Loizou, J.I., and Nik-Zainal, S. (2018). Validating the concept of mutational signatures with isogenic cell models. *Nat. Commun.* 9, 1744–1816. <https://doi.org/10.1038/s41467-018-04052-8>.
  43. Ba, W., Yan, Y., Reijnders, M.R.F., Schuurs-Hoeijmakers, J.H.M., Feenstra, I., Bongers, E.M.H.F., Bosch, D.G.M., De Leeuw, N., Pfundt, R., Gilissen, C., et al. (2016). TRIO loss of function is associated with mild intellectual disability and affects dendritic branching and synapse function. *Hum. Mol. Genet.* 25, 892–902. <https://doi.org/10.1093/hmg/ddv618>.
  44. Jin, Z.B., Wu, J., Huang, X.F., Feng, C.Y., Cai, X.B., Mao, J.Y., Xiang, L., Wu, K.C., Xiao, X., Kloss, B.A., et al. (2017). Trio-based exome sequencing arrests de novo mutations in early-onset high myopia. *Proc. Natl. Acad. Sci. U S A* 114, 4219–4224. <https://doi.org/10.1073/pnas.1615970114>.
  45. Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* 46, 944–950. <https://doi.org/10.1038/ng.3050>.
  46. Kratz, C.P., Freycon, C., Maxwell, K.N., Nichols, K.E., Schiffman, J.D., Evans, D.G., Achatz, M.I., Savage, S.A., Weitzel,

- J.N., Garber, J.E., et al. (2021). Analysis of the Li-fraumeni spectrum based on an international germline TP53 variant data set: an international agency for research on cancer TP53 database analysis. *JAMA Oncol.* 7, 1800. <https://doi.org/10.1001/jamaoncol.2021.4398>.
47. Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., et al. (2020). A structural variation reference for medical and population genetics. *Nature* 581, 444–451. <https://doi.org/10.1038/s41586-020-2287-8>.
48. Fahed, A.C., Wang, M., Homburger, J.R., Patel, A.P., Bick, A.G., Neben, C.L., Lai, C., Brockman, D., Philippakis, A., Ellinor, P.T., et al. (2020). Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat. Commun.* 11, 3635–3639. <https://doi.org/10.1038/s41467-020-17374-3>.
49. Wand, H., Lambert, S.A., Tamburro, C., Iacocca, M.A., O’Sullivan, J.W., Sillari, C., Kullo, I.J., Rowley, R., Dron, J.S., Brockman, D., et al. (2021). Improving reporting standards for polygenic scores in risk prediction studies. *Nature* 591, 211–219. <https://doi.org/10.1038/s41586-021-03243-6>.
50. Sud, A., Hodgson, K., Bloch, G., and Upshur, R. (2021). A conceptual framework for continuing medical education and population Health. *Oncol* 5, 1–15. <https://doi.org/10.1080/10401334.2021.1950540>.
51. Seplyarskiy, V.B., Soldatov, R.A., Koch, E., McGinty, R.J., Goldmann, J.M., Hernandez, R.D., Barnes, K., Correa, A., Burchard, E.G., Ellinor, P.T., et al.; NHLBI Trans-Omics for Precision Medicine TOPMed Consortium<sup>†</sup>; and TOPMed Population Genetics Working Group (2021). Population sequencing data reveal a compendium of mutational processes in the human germ line. *Science* 373, 1030–1035. <https://doi.org/10.1126/science.aba7408>.
52. Mailman, M.D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Boutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L., et al. (2007). The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* 39, 1181–1186. <https://doi.org/10.1038/ng1007-1181>.
53. Crompton, B.D., Stewart, C., Taylor-Weiner, A., Alexe, G., Kurek, K.C., Calicchio, M.L., Kiezun, A., Carter, S.L., Shukla, S.A., Mehta, S.S., et al. (2014). The genomic landscape of pediatric ewing sarcoma. *Cancer Discov.* 4, 1341. <https://doi.org/10.1158/2159-8290.cd-13-1037>.
54. Perry, J.A., Kiezun, A., Tonzi, P., Van Allen, E.M., Carter, S.L., Baca, S.C., Cowley, G.S., Bhatt, A.S., Rheinbay, E., Pedamallu, C.S., et al. (2014). Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. *Proc. Natl. Acad. Sci. U S A* 111. 201419260. <https://doi.org/10.1073/pnas.1419260111>.
55. Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L.D., and Ferretti, V. (2019). The international cancer genome Consortium data portal. *Nat. Biotechnol.* 37, 367–369. <https://doi.org/10.1038/s41587-019-0055-9>.
56. Heath, A.P., Taylor, D.M., Zhu, Y., Raman, P., Lilly, J., Storm, P., Waanders, A.J., Ferretti, V., Yung, C., Mattioni, M., et al. (2019). Abstract 2464: Gabriella Miller Kids First Data Resource Center: harmonizing clinical and genomic data to support childhood cancer and structural birth defect research. *Cancer Res.* 79, 2464. <https://doi.org/10.1158/1538-7445.am2019-2464>.
57. Auton, A., Abecasis, G.R., Brooks, L.D., Altshuler, D.M., Durbin, R.M., Garrison, E.P., Bentley, D.R., Kang, H.M., Chakravarti, A., Korbel, J.O., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. <https://doi.org/10.1038/nature15393>.