CANCER GENETICS

original reports

Pathogenic Germline Variants in Cancer Susceptibility Genes in Children and Young Adults With Rhabdomyosarcoma

Jung Kim, PhD[1]; Nicholas Light, MSc[2,3]; Vallijah Subasri, BS[2,4,5]; Erin L. Young, PhD[6]; Talia Wegman-Ostrosky, MD, PhD[1,7]; Donald A. Barkauskas, PhD[8,9]; David Hall, MS[8]; Philip J. Lupo, PhD, MPH[10]; Rajesh Patidar, MS[11]; Luke D. Maese, DO[6]; Kristine Jones, BS[13]; Mingyi Wang, PhD[13]; Cancer Genome Research Laboratory[13]; Sean V. Tavtigian, PhD[12]; Dongjing Wu, MS[13]; Adam Shlien, PhD[16,17]; Frank Telfer, BSc[2,4]; Anna Goldenberg, PhD[2,5,18]; Stephen X. Skapek, MD[19]; Jun S. Wei, PhD[11]; Xinyu Wen, MS[11]; Daniel Catchpoole, PhD[20]; Douglas S. Hawkins, MD[21]; Joshua D. Schiffman, MD[6,12]; Javed Khan, MD[11]; David Malkin, MD[2,3,4,22]; and Douglas R. Stewart, MD[1]

abstract

PURPOSE Rhabdomyosarcoma (RMS) is the most common pediatric soft-tissue sarcoma and accounts for 3% of all pediatric cancer. In this study, we investigated germline sequence and structural variation in a broad set of genes in two large, independent RMS cohorts.

MATERIALS AND METHODS Genome sequencing of the discovery cohort (n = 273) and exome sequencing of the secondary cohort (n = 121) were conducted on germline DNA. Analyses were performed on 130 cancer susceptibility genes (CSG). Pathogenic or likely pathogenic (P/LP) variants were predicted using the American College of Medical Genetics and Genomics (ACMG) criteria. Structural variation and survival analyses were performed on the discovery cohort.

RESULTS We found that 6.6%-7.7% of patients with RMS harbored P/LP variants in dominant-acting CSG. An additional approximately 1% have structural variants (ATM, CDKN1C) in CSGs. CSG variants did not influence survival, although there was a significant correlation with an earlier age of tumor onset. There was a nonsignificant excess of P/LP variants in dominant inheritance genes in the patients with FOXO1 fusion–negative RMS patients versus the patients with FOXO1 fusion–positive RMS. We identified pathogenic germline variants in CSGs previously (TP53, NF1, DICER1, mismatch repair genes), rarely (BRCA2, CBL, CHEK2, SMARCA4), or never (FGFR4) reported in RMS. Numerous genes (TP53, BRCA2, mismatch repair) were on the ACMG Secondary Findings 2.0 list.

CONCLUSION In two cohorts of patients with RMS, we identified pathogenic germline variants for which gene-specific therapies and surveillance guidelines may be beneficial. In families with a proband with an RMS-risk P/LP variant, genetic counseling and cascade testing should be considered, especially for ACMG Secondary Findings genes and/or with gene-specific surveillance guidelines.

JCO Precis Oncol 5:75-87. © 2021 by American Society of Clinical Oncology

INTRODUCTION

Rhabdomyosarcoma (RMS) accounts for 3% of all pediatric cancer and is the most common soft-tissue tumor in childhood and adolescence.[1] The WHO classification scheme divides RMS into four subtypes: alveolar (ARMS), embryonal (ERMS), pleomorphic, and sclerosing or spindle cell.[2] ERMS accounts for 70% of all RMS and is associated with heterozygous loss of 11p15.5.[3] ERMS typically presents in children under 10 years of age, most commonly in the head and neck region and genitourinary system. ARMS is most commonly driven by the PAX3-FOXO1 or PAX7-FOXO1 gene fusions, termed fusion-positive RMS. The incidence of ARMS is similar across age groups; however, compared with ERMS, the median age at diagnosis is higher.[4,5] The most common site for ARMS is in the deep tissues of the extremities.[6] Five-year event-free survival (EFS) for RMS is around 53%-67% for children younger than 15 years and 30%-51% for children older than 15 years.[7] A multimodal approach

to treatment is used across risk groups (low, intermediate, high), which includes surgery, radiotherapy, and chemotherapy, typically vincristine, dactinomycin, and cyclophosphamide.[8] Risk groups are determined by the TNM stage, clinical group, and fusion status; approximately 50% of the patients are classified as intermediate-risk.[9] RMS survivors are at risk for long-term morbidities, including treatment-related secondary malignancies.[10]

Pathogenic germline variants are observed in 10%-12% of children with cancer.[11-13] One study found that approximately 4% of patients with advanced adult cancers harbored pathogenic germline variants that suggested targets for therapy.[14] RMS is associated with numerous hereditary cancer disorders, including Li-Fraumeni, Lynch, neurofibromatosis type 1, Noonan, Costello, DICER1, and Beckwith-Wiedemann syndromes. In addition, several studies have identified RMS-associated somatic variation, including NRAS, KRAS, HRAS, FGFR4, PIK3CA, CTNNB1, FBXW7,

CONTEXT

Key Objective

To determine the prevalence of germline pathogenic or likely pathogenic (P/LP) variants in known cancer susceptibility genes (CSG) in children and young adults with rhabdomyosarcoma (RMS).

Knowledge Generated

In patients with RMS, approximately 7% harbor germline P/LP variants and approximately 1% harbor structural variants in dominant-acting CSG, including *TP53*, *NF1*, *DICER1*, mismatch repair genes, *BRCA2*, *CBL*, *CHEK2*, and *SMARCA4*. RMS patients with P/LP variants in CSG had an earlier age of tumor onset compared with patients without P/LP variants, although these variants did not influence overall survival. Numerous genes were on the American College of Medical Genetics and Genomics (ACMG) Secondary Findings 2.0 list.

Relevance

Children with RMS should have germline sequencing. Those who have a germline P/LP variant in a CSG should have genetic counseling and cascade testing for other family members, especially for ACMG Secondary Findings genes and/or those with gene-specific surveillance guidelines.

and *BCOR* with Tyr kinase/RAS/PIK3CA axis mutations, and *MYOD1* mutations.[3,15]

The characterization of RMS-associated germline genetic variants may influence the selection of therapeutic agents (eg, PARP [poly(ADP-ribose) polymerase] inhibitors for DNA repair defects), enable cascade genetic testing of at-risk family members, facilitate tumor surveillance for those at risk, and increase understanding of RMS etiology. To identify germline variants, comprehensive analysis of large numbers of individuals with clinical- and molecular-annotated RMS is required. Thus, we investigated germline sequence and structural variation in a broad set of genes in two large and independent RMS cohorts: (1) the discovery cohort, derived from a clinical trial open to patients with intermediate-risk RMS; and (2) the secondary cohort, derived from an independent group of patients with mixed clinical risk and treatment.

MATERIALS AND METHODS

Patients and Ethics Statement

For the discovery cohort, patient DNA samples were collected from the Children's Oncology Group (COG) clinical therapeutic trial ARST0531 investigating intermediate-risk RMS.[8] Consent for banking and future use of germline material was obtained through the linked COG study D9902 (NCT00919269), a companion biobanking protocol to COG ARST0531. For the secondary cohort, patient DNA samples were collected from the Cooperative Human Tissue Network (n = 31), COG (n = 69, D9902, and D9602; NCT00002995), Children's Hospital at Westmead (n = 19), and NCI clinical trial 10-C-0086 (n = 2). Histologic diagnosis and clinical information were compiled by each source, and histology was reviewed when available. Written consent was obtained from all participants (or their parent or guardian). All protocols underwent institutional review board review and approval.

Sequencing, Filtering, and Annotation

Exome (ES) and genome sequencing (GS) was processed and mapped using previously published methods.[16] ANNOVAR[17] was used to annotate gene, population allele frequencies, and in silico pathogenicity. ClinVar classification was annotated using clinical laboratories meeting minimum requirements for data sharing to support quality assurance (badged lab) (2018-06-04 version).[18] InterVar classification was annotated using python version 0.1.7 20180118.[19] All variants were filtered using publicly available databases (Exome Sequencing Project, 1000 Genomes Project, Exome Aggregation Database without The Cancer Genome Atlas [TCGA], Genome Aggregation Database without TCGA) where any population allele frequency <1%. Details are in the Appendix Supplementary Methods.

Cancer Susceptibility Genes

Analysis focused on germline variants in 130 cancer susceptibility genes (CSG) from the study of Rahman[20] and published RMS-associated genes[21] (Data Supplement, online only). These candidate genes include those with autosomal-dominant (n = 76), autosomal-recessive (n = 28), both autosomal-dominant and autosomal-recessive inheritance (n = 16), X-linked (n = 5), Y-linked (n = 1), and unknown (n = 4) modes of inheritance.

Automated and Manual Review of Variation

The Appendix Figure A1 shows the scheme by which variant pathogenicity was determined. Panel sequencing was performed on a subset of genes. Details are given in the Appendix Supplementary Methods.

Structural Variant Analysis

Germline structural variant (SV) analysis was performed on the discovery cohort using DELLY v0.7.7[22] and Manta v1.0.3[23] for the discovery of deletions, duplications, and inversions, and

MELT v2.0.5[24] for the discovery of mobile element insertions (MEIs) in 130 CSG (Appendix Supplementary Methods).

PCR Validation of Structural Variation

A series of polymerase chain reactions (PCR) were performed to validate breakpoints identified by bioinformatic means; the PCR primers used and experimental details are listed in the Appendix Supplementary Methods.

Statistical Analysis

A total of two predictor variables (pathogenic or likely pathogenic [P/LP] variants in 130 CSG with autosomal dominant (AD) and autosomal dominant and autosomal recessive (AD/AR) genes [102 genes]) were analyzed for association in the 256 RMS patients with confirmed intermediate-risk features and for which EFS and overall survival (OS) data were available in the discovery cohort. Each predictor variable was tested for univariate association with EFS and OS using log-rank tests and for association within multivariate Cox proportional hazard regressions. The *P*- value presented for each of the univariate models is the *P*-value from the log-rank test. Details are in the Appendix Supplementary Methods.

RESULTS

Patient Demographics

Table 1 lists the demographic and clinical data for the discovery (n = 273 germline GS) and secondary (n = 121 germline ES) cohorts. The distributions of sex, histology, and fusion status in the discovery cohort are similar to what has been reported in the literature for intermediate RMS. The distribution of histology and fusion status in the secondary cohort reflects the heterogeneity of risk in that group.

Frequency of P/LP Variants in Dominant and Dominant or Recessive CSG was 7.7% in the Discovery Cohort and 6.6% in the Secondary Cohort

In 43 patients in the discovery cohort, there were 44 unique P/LP variants (allele count = 58) in 23 genes; in 16 patients in the secondary cohort, there were 17 unique variants (allele count = 21) in 13 genes (Table 2). Using genes with dominant and dominant or recessive inheritance (102 CSGs), the frequency of P/LP variants was 7.7% in the discovery cohort and 6.6% in the secondary cohort (Table 3). Among *FOXO1* fusion–negative tumors, the frequency of P/LP variants (by-person) in all 130 CSG was 15.5% in the discovery cohort and 12.5% in the secondary cohort (Table 2); using the 102 genes with only dominant and dominant or recessive inheritance, the frequency of P/LP variants (by-person) was 9.8% in the discovery cohort and 8.3% in the secondary cohort (Table 3).

Germline P/LP Variants Observed in Known and Novel RMS-Associated Genes

At the variant-specific level, there were seven heterozygous variants (Table 4) observed in both cohorts in genes with dominant (*TP53*) and recessive (*MUTYH*, *GBA*, *SBDS*,

TABLE 1. Demographics of Participants Enrolled in the Discovery and Secondary Rhabdomyosarcoma Cohorts

Number of patients (%)	Discovery (n = 273)	Secondary (n = 121)
Histology		
Alveolar	112 (41.0%)	49 (40.5%)
Embryonal	136 (49.8%)	56 (46.3%)
Mixed alveolar/embryonal	0 (0%)	6 (5.0%)
Embryonal/botryoid	0 (0%)	1 (0.8%)
Botryoid	13 (4.8%)	0 (0%)
Other, specify	7 (2.6%)	5 (4.1%)
Embryonal ectomesenchymoma	1 (0.4%)	0 (0%)
Spindle cell	3 (1.0%)	0 (0%)
Unknown	1 (0.4%)	4 (3.4%)
Sex		
Male	146 (53.5%)	65 (53.7%)
Female	127 (46.5%)	56 (46.3%)
FOXO1 fusion status (alveolar only)		
Fusion-positive	62 (55.4%)	33 (67.3%)
Fusion-negative	18 (16.1%)	5 (10.2%)
Unknown	32 (28.6%)	11 (22.4%)
Risk group		
Low-risk	3 (1.1%)	14 (11.6%)
Intermediate-risk	256 (93.8%)	32 (26.4%)
High-risk	0 (0.0%)	23 (19.0%)
Unknown	14 (5.1%)	52 (43.0%)
Race		
Black or African American	29 (10.6%)	1 (0.8%)
White	202 (74.0%)	80 (66.1%)
Unknown or not reported	31 (11.4%)	0 (0%)
Asian	8 (2.9%)	3 (2.5%)
American Indian/Alaska Native	3 (1.1%)	0 (0%)
Admix	0 (0%)	37 (30.6%)
Age at enrollment (range), years	7.5 (0-25)	7.6 (0-20)

ERCC2) patterns of inheritance. Biallelic *ERCC2* variants (p.A717G and p.L461V) were observed in one person in both the discovery and secondary cohorts; phase could not be determined but is likely *in cis*.[25-27] At the gene-specific level, P/LP variants were observed in eight genes in both cohorts, including *DICER1*, *TP53* (dominant inheritance), *ATM*, *COL7A1* (dominant and recessive inheritance), *ERCC2*, *GBA*, *MUTYH*, and *SBDS* (recessive inheritance). At the pathway-specific level, variation in Lynch syndrome genes (*MSH2* and *MSH6*) was found in both cohorts.

In addition to observing P/LP variants in established, dominant RMS-associated genes (*TP53*, *DICER1*, *NF1*, *MSH6*, *MSH2*), we observed P/LP variants in dominant-acting genes in which pathogenic germline variation has never (or rarely) been reported in RMS, including *BRCA2*, *CBL*, *CHEK2*, *FH*, *RET*, and

TABLE 2. Frequency of Pathogenic or Likely Pathogenic Germline Variants in All 130 Cancer Susceptibility Genes

	Discovery (n = 273) Unique Count (Allele Count)			Secondary (n = 121) Unique Count (Allele Count)		
	All (n = 273)	**Confirmed Fusion-Positive (n = 67)**	**Fusion-Negative (n = 174)**	**All (n = 121)**	**Confirmed Fusion-Positive (n = 38)**	**Fusion-Negative (n = 72)**
Pathogenic	32 (43)	10 (12)	20 (26)	14 (18)	6 (7)	10 (11)
Likely pathogenic	12 (15)	2 (3)	10 (11)	3 (3)	—	3 (3)
Frequency-by-allele (95% CI)	21.3% (16.6 to 26.6)	22.4% (13.1 to 34.2)	21.3% (15.4 to 28.1)	17.4% (11.1 to 25.3)	18.4% (7.7 to 34.3)	19.4% (11.1 to 30.5)
Frequency-by-person (95% CI)	15.8% (11.6 to 20.6)	16.4% (8.5 to 27.5)	15.5% (10.5 to 21.8)	13.2% (7.8 to 20.6)	15.8% (6.0 to 31.3)	12.5% (5.6 to 22.4)

NOTE: Fusion-positive tumors reflect laboratory-confirmed fusion status. Fusion-negative tumors are alveolar rhabdomyosarcoma with confirmed fusion-negative status and any nonalveolar rhabdomyosarcoma patients that were not tested. For discovery cohort, there were a total of 67 fusion-positive tumors, reflecting 62 alveolar tumors (Table 1), plus 5 embryonal rhabdomyosarcoma and other/NOS tumors.

SMARCA4 (Fig 1, Data Supplement). For BRCA2, we observed one P/LP allele in a single individual and two P/LP alleles in a second individual in the discovery cohort. Phase could not be determined for the biallelic BRCA2 variants because of genomic distance. A missense variant (p.L318P) was observed in FGFR4, a gene somatically mutated in RMS,[21] but to date without recognized germline phenotypic consequence (predicted LP by InterVar). A custom-built multigene panel using germline DNA from patients in the discovery cohort verified 100% (9/9) of the tested P/LP variants.

P/LP Germline Variants in Dominant Genes in Participants with FOXO1 Fusion–Negative Tumors

In the discovery cohort, 67/273 (25%) of the patients had a confirmed fusion-positive tumor; of those, 2/67 (3.0%) harbored a P/LP variant in genes with a dominant mode of inheritance, whereas 17/174 (9.8%) patients with a fusion-negative tumor had a dominant P/LP variant. This difference trended toward significance (Fisher's exact test, $P = .08$). In the secondary cohort, 38/121 (31%) of the patients had a confirmed fusion-positive tumor; of those, 2/38 (5.3%) harbored P/LP variants in genes with a dominant mode of inheritance, whereas 6/72 (8.3%) patients

with a fusion-negative tumor had a dominant P/LP variant (Fisher's exact test, $P = .56$).

Histology and Demographics of Patients With Germline P/LP Variants in CSG

We used OncoPrint to visualize P/LP variants and clinical data, including age at enrollment, stage, histology, and fusion status (Fig 1, Data Supplement). We have very limited demographic data for the secondary cohort compared with discovery cohort (Data Supplement). Therefore, only sex was tested for the secondary cohort. The age at enrollment for patients harboring any CSG P/LP variants (mean = 6.8 years) was younger than those without any CSG P/LP variants (mean = 7.6 years) but was not statistically significant (Mann-Whitney U test, $P = .21$). However, if restricted to genes with dominant or dominant or recessive inheritance, patients with P/LP variants were significantly younger (mean = 4.4 years) compared with patients without P/LP variants (mean = 7.7 years; Mann-Whitney U test, $P = .002$; Fig 2). There were no significant differences in the sex of the patients in regard to the risk of P/LP in both cohorts (Fisher's exact test, $P = .87$ and .59).

TABLE 3. Frequency of Pathogenic or Likely Pathogenic Germline Variants in 102 Dominant and Dominant or Recessive Cancer Susceptibility Genes

	Discovery (n = 273) Unique Count (Allele Count)			Secondary (n = 121) Unique Count (Allele Count)		
	All (n = 273)	**Confirmed Fusion-Positive (n = 67)**	**Fusion-Negative (n = 174)**	**All (n = 121)**	**Confirmed Fusion-Positive (n = 38)**	**Fusion-Negative (n = 72)**
Pathogenic	18 (18)	2 (2)	15 (15)	7 (7)	2 (2)	5 (5)
Likely pathogenic	3 (3)	—	3 (3)	1 (1)	—	1 (1)
Frequency-by-allele (95% CI)	7.7% (4.8 to 11.5)	3.0% (0.4 to 10.4)	10.3% (6.3 to 15.9)	6.6% (2.9 to 12.6)	5.3% (0.6 to 17.8)	8.3% (3.1 to 17.3)
Frequency-by-person (95% CI)	7.7% (4.8 to 11.5)	3.0% (0.4 to 10.4)	9.8% (5.8 to 15.2)	6.6% (2.9 to 12.6)	5.3% (0.6 to 17.8)	8.3% (3.1 to 17.3)

TABLE 4. Pathogenic or Likely Pathogenic Variants Found in Both the Discovery and Secondary Cohorts

Gene	c.HGVS: p. HGVS
MUTYH	c.G1187A:p.G396D
MUTYH	c.A536G:p.Y179C
GBA	c.T1448C:p.L483P
SBDS	c.258+2T>C
TP53	c.G347A:p.R116Q
ERCC2	c.C2150G:p.A717G
ERCC2	c.C1381G:p.L461V

NOTE: Alleles were observed just once in each cohort. Both ERCC2 variants were observed in a single individual in each cohort. Phase could not be determined with certainty, but published data suggests the variants are likely in *cis*.

Heterozygous Deletions in *CDKN1C* and *ATM* Identified in Two Patients

Analysis of structural variation identified two individuals with PCR-confirmed deletions in *ATM* and *CDKN1C* (Appendix Fig A2) that were absent from noncancer cohorts (gnomAD and 1000 Genomes Project). One subject with an ERMS harbored a deletion spanning exon 1 of *CDKN1C*, a tumor-suppressor gene associated with Beckwith-Wiedemann syndrome, a congenital overgrowth disorder characterized by an increased susceptibility to pediatric cancer, including RMS. A second individual with ERMS harbored a deletion in the kinase domain of the DNA-damage response tumor suppressor, *ATM*, thus likely altering stability and impairing protein kinase activity. Neither individual harbored concomitant sequence P/LP variants in any of the 130 analyzed CSG. No pathogenic duplications, inversions, or MEIs were discovered in CSGs in any of the patients. Overall, we observed structural variation affecting the exons of CSGs in 2/273 (0.7%) of children with RMS.

No Significant Differences in OS or EFS in Participants With P/LP Variants

There was no significant difference in OS or EFS between P/LP carriers versus non-P/LP carriers (Appendix Fig A3) for either univariate or multivariate analysis except univariate analysis with age at enrollment for 102 AD, AD/AR genes (*P*-value = .01).

FIG 1. Oncoprint of pathogenic and likely pathogenic variants in 130 cancer susceptibility genes in the (A) discovery cohort (B) secondary cohort, sorted by histology type.
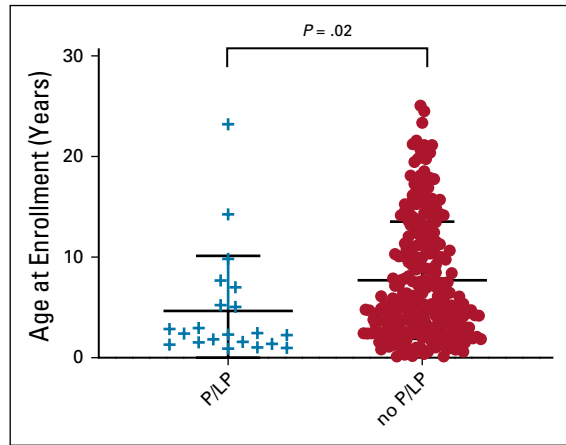
FIG 2. Box-and-whisker plot comparing age at enrollment in dominant and dominant or recessive CSG in P/LP variant carriers *v* no P/LP variant carriers in discovery RMS patients. CSG, cancer susceptibility genes; P/LP, pathogenic or likely pathogenic; RMS, rhabdomyosarcoma.

DISCUSSION

We used a conservative, clinically based approach to classify germline variation in a set of predefined CSG in discovery (n = 273, predominately intermediate-risk) and secondary (n = 121, unselected risk group) cohorts of children and young adults with RMS. Our study is notable for the large size of two independent and fully sequenced cohorts, a comprehensive analysis of genome-wide SVs of the discovery cohort, and survival analyses of patients harboring germline pathogenic variants. We note that in a separate, unselected cohort of 615 RMS pediatric patients, 7.3% carried a cancer predisposition variant in a known CSG, in comparison to 1.4% in population-based controls.[28] We identified pathogenic germline variants in numerous CSG that have been previously (*TP53*, *NF1*, *DICER1*, mismatch repair genes), rarely (*BRCA2*, *CBL*, *CHEK2*, *SMARCA4*), or never (*FGFR4*) reported in children and young adults with RMS. The frequency and type of P/LP variants we found are comparable to another recent, large exome sequencing study of germline risk in RMS.[28] We identified pathogenic structural variation (*ATM*, *CDKN1C*) in two individuals and estimate the frequency of such variation in the RMS population. Our work found an excess of dominant-acting variants in individuals with fusion-negative tumors, which has implications for the design of future germline discovery studies. Our work can inform clinical care: In two cohorts of patients with RMS, we identified pathogenic variants for which gene-specific therapies and surveillance guidelines may be beneficial. In addition, in families with P/LP variants in RMS-risk genes, genetic counseling and cascade testing should be considered for all members, regardless of age.

Using the American College of Medical Genetics and Genomics (ACMG) or Association for Molecular Pathology

criteria, the most common pathogenic germline variants we found were in *TP53*, *NF1*, *DICER1*, and the mismatch repair pathway genes. These variants were observed exclusively in individuals with nonalveolar histology. In both the discovery and secondary cohorts, we observed a nonsignificant excess of P/LP alleles in genes with dominant inheritance in individuals with *FOXO1* fusion–negative tumors. These two complementary observations highlight the potency of *FOXO1* fusion genes and suggest that the effect of germline variation in *FOXO1* fusion–positive tumors in the pathogenesis of RMS may be of secondary importance. These observations have consequences in the design of studies to identify germline variation that increases the risk for RMS. To be adequately powered, studies investigating the effects of dominant-acting variation should emphasize the recruitment of individuals with *FOXO1* fusion–negative tumors. The role of germline variation in the pathogenesis of *FOXO1* fusion–positive tumors (if any) also would require adequately powered exome studies with appropriate controls.

This study uncovered pathogenic variants in genes not previously linked to RMS susceptibility. In both the discovery and secondary cohorts, a single individual with biallelic *ERCC2* variants was observed. Previously published work shows that the two alleles, p.A717G and p.L461V, are found together *in cis* and create a complex allele that is associated with decreased DNA repair function.[26] Overall, for recessive genes, heterozygous alleles in *ERCC2*, *GBA*, *MUTYH*, *SBDS* were found in both cohorts; however, given the lack of matched controls, the meaning of these observations is uncertain. In the discovery cohort, we identified P/LP variants in *CBL* (associated with a Noonan-like disorder) and *SMARCA4* (associated with risk for rhabdoid tumors, a histologically distinct entity from RMS). In the secondary cohort, we found P/LP variants in *RET* (multiple endocrine neoplasia type 2), and *FH* (hereditary leiomyomatosis and renal cell cancer). With the exception of *CBL*-associated Noonan,[29] these other genes have not been previously linked to RMS. Our findings, if replicated, expand the phenotypic consequences of P/LP variants in *ERCC2*, *CBL*, *SMARC4*, *RET*, and *FH* and should inform revision of genetic counseling and clinical care for individuals harboring these variants. A rare variant in *FGFR4* (somatically mutated in RMS[21]) was classified as LP. However, since there is no known human, phenotypic consequence arising from *FGFR4* germline variants and since this was an InterVar-based classification, this finding should be interpreted with caution but merits follow-up.

Pediatric cancer, including RMS, has been observed in families with germline pathogenic *BRCA1* and *BRCA2* variants.[30,31] In pediatric pan-cancer studies, pathogenic variants in *BRCA2* were observed in individuals with RMS in two different reports (1/43 [2.3%][12]; 1/21 [4.8%][13]). In our data, we observed P/LP variants in *BRCA2* in 1% of *FOXO1*

fusion–negative RMS in the discovery cohort, a rate comparable to other similar-sized studies.[28] We observed P/LP variants in other adult cancer genes, including sequence (in both cohorts) and structural variation in *ATM*. Lastly, we observed a single participant with a P/LP variant in *CHEK2*, a gene with wide pleiotropy and recognized risk for a variety of adult-onset cancers, including sarcomas.[32] Our observations of RMS susceptibility from germline pathogenic variants in *ATM, BRCA2, CHEK2, MSH2,* and *MSH6* blur the historic distinction in risk between adult and pediatric cancer genes, especially given the statistically significant earlier age of enrollment of children with RMS and dominant-acting P/LP variants. Some of these genes (*TP53, BRCA2,* mismatch repair) are on the ACMG Secondary Findings 2.0 list.[33] Larger studies are needed to quantify the risk of RMS from germline pathogenic variants. For now, in families with a proband with an RMS-risk P/LP variant, genetic counseling and cascade testing should be considered, especially for genes with specific surveillance guidelines (eg, National Comprehensive Care Network [NCCN], *DICER1*[34]) or on the ACMG Secondary Findings list.

Strengths of our study include its large size, independent cohorts broadly representative of North American pediatric cancer populations, clinically based classification of pathogenicity, and availability of structural variation analysis in the discovery cohort. Limitations include a lack of family history data and, in the secondary cohort, more limited clinical information. This study of predominantly intermediate-risk and unselected RMS may not be generalizable to all RMS-risk groups.

In this study, we found that 8%-10% of children and young adults with predominately intermediate-risk, fusion-negative RMS harbored P/LP variants in dominant-acting CSG. An additional approximately 1% of patients with RMS have structural germline variants in CSGs. There are multiple clinical and research consequences arising from our findings. To be adequately powered, studies investigating the effects of variants in dominant-acting CSGs should emphasize the recruitment of individuals with *FOXO1* fusion–negative tumors. Clinically, the identification of such germline variation for the individual could be critically important to fulfill eligibility criteria for clinical trials to guide targeted treatments (eg, use of MEK [for RAS pathway variants] or PARP inhibitors for DNA repair variants) and permits gene-specific imaging and/or laboratory and imaging surveillance available now for many CSG. The identification of germline variation in patients with RMS is critically important as it would prompt genetic counseling and cascade genetic testing to identify other family members, whose risk of RMS (or numerous other cancers, as with *DICER1, MSH2, MSH6, NF1, RET, TP53*) may then be prospectively managed with NCCN or other published guidelines.[34] This is especially true for genes on the ACMG Secondary Findings list.[33] In children and young adults with RMS without a pathogenic germline P/LP variant, detection of germline structural variation in RMS-risk genes should be considered. More speculatively, we identified individuals with RMS and pathogenic germline variants in *ERCC2, FGFR4, FH, RET,* and *SMARCA4* which, if replicated, may contribute to further understanding of RMS etiology and opportunities for intervention.

AFFILIATIONS

[1]Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD

[2]Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, ON, Canada

[3]Institute of Medical Science, University of Toronto, Toronto, ON, Canada

[4]Department of Medical Biophysics, University of Toronto, ON, Canada

[5]Vector Institute of Artificial Intelligence, Toronto, ON, Canada

[6]Department of Pediatrics, University of Utah, Salt Lake City, UT

[7]Basic Research Subdirection, Instituto Nacional de Cancerología (INCan), Mexico City, Mexico

[8]QuadW-COG Childhood Sarcoma Biostatistics and Annotation Office, Children's Oncology Group, Monrovia, CA

[9]Department of Preventive Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA

[10]Department of Pediatrics, Hematology-Oncology Section, Baylor College of Medicine, Houston, TX

[11]Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD

[12]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT

[13]Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD

[14]Department of Oncological Sciences, University of Utah School of Medicine, Salt Lake City, UT

[15]Huntsman Cancer Institute at the University of Utah, Salt Lake City, UT

[16]Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada

[17]Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, ON, Canada

[18]Department of Computer Science, University of Toronto, Toronto, ON, Canada

[19]University of Texas Southwestern, Dallas, TX

[20]The Tumour Bank, Children's Cancer Research Unit, Kids Research Institute, The Children's Hospital at Westmead, Westmead, NSW, Australia

[21]Division of Hematology/Oncology, Seattle Children's Hospital, University of Washington, Fred Hutchinson Cancer Research Center, Seattle, WA

[22]Division of Hematology-Oncology, The Hospital for Sick Children, Department of Pediatrics, University of Toronto, Toronto, ON, Canada

CORRESPONDING AUTHOR

Javed Khan, MD, Center for Cancer Research, National Cancer Institute, Building 37, Room 2016B, Bethesda, MD 20892; e-mail: khanjav@nih.gov.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/po/authors/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Talia Wegman-Ostrosky
Travel, Accommodations, Expenses: Pfizer

Donald A. Barkauskas
Employment: Genentech
Stock and Other Ownership Interests: Genentech
Patents, Royalties, Other Intellectual Property: U.S. patent based on Ph.D. research in glioblastoma

Luke D. Maese
Honoraria: Jazz Pharmaceuticals
Consulting or Advisory Role: Jazz Pharmaceuticals

Sean V. Tavtigian
Patents, Royalties, Other Intellectual Property: I hold patents on BRCA1, BRCA2, and PTEN. However, the claims in these patents related to genetic testing have been overturned by the US Supreme Court. Continuing claims are irrelevant to the paper. My royalty stream ended more than 2 years ago.

Anna Goldenberg
Research Funding: 4UandMe

Douglas S. Hawkins
Research Funding: Loxo, Bristol-Myers Squibb, Merck Sharp Dohme, Bayer, Lilly, Eisai, Amgen, Seattle Genetics, Incyte, Jazz Pharmaceuticals
Travel, Accommodations, Expenses: Loxo, Bayer, Celgene, AstraZeneca

Joshua D. Schiffman
Employment: PEEL Therapeutics, Inc.
Leadership: PEEL Therapeutics, Inc.
Stock and Other Ownership Interests: ItRunsInMyFamily.com, PEEL Therapeutics, Inc.
Honoraria: Affymetrix
Consulting or Advisory Role: N-of-One, Fabric Genomics

Javed Khan
Research Funding: Lentigen
Patents, Royalties, Other Intellectual Property: Monoclonal antibody-based therapeutics targeting fibroblast growth factor receptor 4 (FGFR4) for potential treatment of human cancers expressing FGFR4

David Malkin
Consulting or Advisory Role: Bayer, Canada

Douglas R. Stewart
Employment: Genome Medical, LLC

No other potential conflicts of interest were reported.

AUTHOR CONTRIBUTIONS

Conception and design: Jung Kim, Nicholas Light, Philip J. Lupo, Sean V. Tavtigian, Adam Shlien, Anna Goldenberg, Douglas S. Hawkins, Joshua D. Schiffman, Javed Khan, David Malkin, Douglas R. Stewart
Financial support: Anna Goldenberg, Joshua D. Schiffman, Javed Khan, David Malkin, Douglas R. Stewart
Provision of study materials or patients: Mingyi Wang, Daniel Catchpoole, Javed Khan
Collection and assembly of data: Nicholas Light, Vallijah Subasri, Erin L. Young, Donald A. Barkauskas, David Hall, Philip J. Lupo, Kristine Jones, Sean V. Tavtigian, Frank Telfer, Jun S. Wei, Xinyu Wen, Daniel Catchpoole, Douglas S. Hawkins, Joshua D. Schiffman, Javed Khan
Data analysis and interpretation: Jung Kim, Nicholas Light, Vallijah Subasri, Erin L. Young, Talia Wegman-Ostrosky, Donald A. Barkauskas, David Hall, Philip J. Lupo, Rajesh Patidar, Luke D. Maese, Mingyi Wang, Dongjing Wu, Frank Telfer, Anna Goldenberg, Stephen X. Skapek, Jun S. Wei, Xinyu Wen, Douglas S. Hawkins, Joshua D. Schiffman, Javed Khan, David Malkin
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

REFERENCES

1. Egas-Bejar D, Huh WW: Rhabdomyosarcoma in adolescent and young adult patients: Current perspectives. Adolesc Health Med Ther 5:115-125, 2014
2. Jo VY, Fletcher CD: WHO classification of soft tissue tumours: An update based on the 2013 (4th) edition. Pathology 46:95-104, 2014
3. Shern JF, Yohe ME, Khan J: Pediatric rhabdomyosarcoma. Crit Rev Oncog 20:227-243, 2015
4. Ognjanovic S, Linabery AM, Charbonneau B, et al: Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975-2005. Cancer 115:4218-4226, 2009
5. Skapek SX, Ferrari A, Gupta AA, et al: Rhabdomyosarcoma. Nat Rev Dis Primers 5:1, 2019
6. Kashi VP, Hatley ME, Galindo RL: Probing for a deeper understanding of rhabdomyosarcoma: Insights from complementary model systems. Nat Rev Cancer 15:426-439, 2015
7. Smith MA, Altekruse SF, Adamson PC, et al: Declining childhood and adolescent cancer mortality. Cancer 120:2497-2506, 2014

8. Hawkins DS, Chi YY, Anderson JR, et al: Addition of vincristine and irinotecan to vincristine, dactinomycin, and cyclophosphamide does not improve outcome for intermediate-risk rhabdomyosarcoma: A report from the Children's Oncology Group. J Clin Oncol 36:2770-2777, 2018

9. Breitfeld PP, Meyer WH: Rhabdomyosarcoma: New windows of opportunity. Oncologist 10:518-527, 2005

10. Zong X, Pole JD, Grundy PE, et al: Second malignant neoplasms after childhood non-central nervous system embryonal tumours in North America: A population-based study. Eur J Cancer 84:173-183, 2017

11. Parsons DW, Roy A, Yang Y, et al: Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors. JAMA Oncol 2:616-624, 2016

12. Zhang J, Walsh MF, Wu G, et al: Germline mutations in predisposition genes in pediatric cancer. N Engl J Med 373:2336-2346, 2015

13. Grobner SN, Worst BC, Weischenfeldt J, et al: The landscape of genomic alterations across childhood cancers. Nature 555:321-327, 2018

14. Mandelker D, Zhang L, Kemel Y, et al: Mutation detection in patients with advanced cancer by universal sequencing of cancer-related genes in tumor and normal DNA vs guideline-based germline testing. JAMA 318:825-835, 2017

15. Chen X, Stewart E, Shelat AA, et al: Targeting oxidative stress in embryonal rhabdomyosarcoma. Cancer Cell 24:710-724, 2013

16. Kim J, Luo W, Wang M, et al: Prevalence of pathogenic/likely pathogenic variants in the 24 cancer genes of the ACMG secondary findings v2.0 list in a large cancer cohort and ethnicity-matched controls. Genome Med 10:99, 2018

17. Wang K, Li M, Hakonarson H: ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164, 2010

18. Landrum MJ, Lee JM, Benson M, et al: ClinVar: Improving access to variant interpretations and supporting evidence. Nucleic Acids Res 46:D1062-D1067, 2018

19. Li Q, Wang K: InterVar: Clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. Am J Hum Genet 100:267-280, 2017

20. Rahman N: Realizing the promise of cancer predisposition genes. Nature 505:302-308, 2014

21. Shern JF, Chen L, Chmielecki J, et al: Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. Cancer Discov 4:216-231, 2014

22. Rausch T, Zichner T, Schlattl A, et al: DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28:i333-i339, 2012

23. Chen X, Schulz-Trieglaff O, Shaw R, et al: Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32:1220-1222, 2016

24. Gardner EJ, Lam VK, Harris DN, et al: The mobile element locator tool (MELT): Population-scale mobile element discovery and biology. Genome Res 27:1916-1929, 2017

25. Takayama K, Salazar EP, Lehmann A, et al: Defects in the DNA repair and transcription gene ERCC2 in the cancer-prone disorder xeroderma pigmentosum group D. Cancer Res 55:5656-5663, 1995

26. Zhou X, Khan SG, Tamura D, et al: Abnormal XPD-induced nuclear receptor transactivation in DNA repair disorders: Trichothiodystrophy and xeroderma pigmentosum. Eur J Hum Genet 21:831-837, 2013

27. Orioli D, Compe E, Nardo T, et al: XPD mutations in trichothiodystrophy hamper collagen VI expression and reveal a role of TFIIH in transcription derepression. Hum Mol Genet 22:1061-1073, 2013

28. Li H, Sisoudlya, SD. Khayat, MM, et al: Prevalence of germline cancer-predisposition variants in pediatric rhabdomyosarcoma: a report from the Children's Oncology Group. JNCI, 2020

29. Ji J, Navid F, Hiemenz MC, et al: Embryonal rhabdomyosarcoma in a patient with a germline CBL pathogenic variant. Cancer Genet 231-232:62-66, 2019

30. Brooks GA, Stopfer JE, Erlichman J, et al: Childhood cancer in families with and without BRCA1 or BRCA2 mutations ascertained at a high-risk breast cancer clinic. Cancer Biol Ther 5:1098-1102, 2006

31. Magnusson S, Borg A, Kristoffersson U, et al: Higher occurrence of childhood cancer in families with germline mutations in BRCA2, MMR and CDKN2A genes. Fam Cancer;7:331-337, 2008

32. Naslund-Koch C, Nordestgaard BG, Bojesen SE: 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, 2016

33. Kalia SS, Adelman K, Bale SJ, et al: Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): A policy statement of the American College of Medical Genetics and Genomics. Genet Med 19:249-255, 2017

34. Schultz KAP, Williams GM, Kamihara J, et al: DICER1 and associated conditions: Identification of at-risk individuals and recommended surveillance strategies. Clin Cancer Res 24:2251-2261, 2018

APPENDIX

APPENDIX. SUPPLEMENTARY METHODS

Exome Sequencing (ES)

DNA preparation. For each sample, 200-ng genomic DNA was purified using Agencourt AMPure XP Reagent (Beckman Coulter Inc, Brea, CA) according to the manufacturer's protocol. An adapter-ligated library was prepared with the KAPA HyperPlus Kit (KAPA Biosystems, Wilmington, MA) using Bioo Scientific NEXTflex DNA Barcoded Adapters (Bioo Scientific, Austin, TX) according to the KAPA-provided protocol.

Prehybridization Ligation Mediated (LM)-Polymerase Chain Reaction (PCR). Genomic DNA sample libraries were amplified prehybridization by ligation-mediated PCR consisting of one reaction containing 20 μL library DNA, 25 μL 2× KAPA HiFi HotStart ReadyMix, and 5 μL 10× Library Amplification Primer Mix (which includes two primers whose sequences are 5′-AATGATACGGCGACCACCGA-3′ and 5′-CAAGCAGAAGACGGCATACGA-3′). PCR cycling conditions were as follows: 98°C for 45 seconds followed by five cycles of 98°C for 15 seconds, 60°C for 30 seconds, and 72°C for 30 seconds. The last step was an extension at 72°C for 1 minute. The reaction was kept at 4°C until further processing. The amplified material was cleaned with Agencourt AMPure XP Reagent (Beckman Coulter Inc, Brea, CA) according to the KAPA-provided protocol. Amplified sample libraries were quantified using Quant-iT PicoGreen dsDNA Reagent (Life Technologies, Carlsbad, CA).

Liquid-Phase Sequence Capture. Before hybridization, amplified sample libraries with unique barcoded adapters were combined in equal amounts into 1.1-μg pools for multiplex sequence capture. Exome sequence capture was performed with NimbleGen's SeqCap EZ Human Exome Library, v3.0, with 64 Mb of exonic sequence targeted (Roche NimbleGen, Inc., Madison, WI). Before hybridization, the following components were added to the 1.1-μg pooled sample library, 4 μL of NEXTflex HE Universal Oligo 1,250 μM (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC-TCTTCCGATCT-3′), 40-μL total 25 μM NEXTflex INV-HE blocking oligos, equal volumes of each blocking oligo complementary to the barcodes in the pool (5′-CAAGCAGAAGACGGCATACGAGAT**X**GTG-ACT GGAGTTCAGACGTGTGCTCTTCCGATCT/C3 Spacer/-3′, where **X** is 8 bases of sequence specific to adapter barcode used for library construction), and 5 μL of 1 mg/mL COT-1 DNA (Invitrogen, Inc, Carlsbad, CA). Samples were dried down by puncturing a hole in the plate seal and processing in an Eppendorf 5301 Vacuum Concentrator (Eppendorf, Hauppauge, NY) set to 60°C for approximately 1 hour. To each dried pool, 7.5 μL of NimbleGen Hybridization Buffer and 3.0 μL of NimbleGen Hybridization Component A were added and placed in a heating block for 10 minutes at 95°C. The mixture was then transferred to 4.5 μL of EZ Exome Probe Library and hybridized at 47°C for 64-72 hours. Washing and recovery of captured DNA were performed as described in the NimbleGen SeqCap EZ Library SR Protocol.

Posthybridization LM-PCR. Pools of captured DNA were amplified by ligation-mediated PCR consisting of one reaction for each pool containing 20 μL captured library DNA, 25 μL 2× KAPA HiFi HotStart ReadyMix, and 5 μL 10× Library Amplification Primer Mix (which includes two primers whose sequences are 5′-AATGA-TACGGCGACCACCGA-3′ and 5′-CAAGCAGAAGACGGCATACGA-3′). PCR cycling conditions were as follows: 98°C for 45 seconds followed by 8 cycles of 98°C for 15 seconds, 60°C for 30 seconds, and 72°C for 30 seconds. The last step was an extension at 72°C for 1 minute. The reaction was kept at 4°C until further processing. The amplified material was cleaned with Agencourt AMPure XP Reagent (Beckman Coulter Inc, Brea, CA) according to the NimbleGen SeqCap EZ Library SR Protocol. Pools of amplified captured DNA were then quantified via Kapa's Library Quantification Kit for Illumina (Kapa Biosystems, Woburn, MA) on the LightCycler 480 (Roche, Indianapolis, IN).

Exome sequencing. The human reference genome and the known gene transcript annotation were downloaded from the UCSC database (http://genome.ucsc.edu/), version hg19 (corresponding to the Genome Reference Consortium assembly GRCh37). Sequencing reads were first trimmed using the Trimmomatic program (v0.32),[1] which marks all low-quality stretches (average quality score < Q15 in a 4-bp sliding window) and reports the longest high-quality stretch of each read. Only read pairs with both ends no shorter than 36 bp were used. Reads were then aligned to the hg19 reference genome using Novoalign software (v3.00.05) (http://www.novocraft.com). Duplicate reads because of either optical or PCR artifacts were removed from further analysis using the MarkDuplicates module of the Picard software (v1.126) (http://picard.sourceforge.net/). Additionally, our analysis used only properly aligned read pairs, in the sense that the two ends of each pair must be mapped to the reference genome in complementary directions and must reflect a reasonable fragment length (300 ± 100 bp). These high-quality alignments for each individual were further refined according to a local realignment strategy around known and novel sites of insertion and deletion polymorphisms using the RealignerTargetCreator and IndelRealigner modules from the Genome Analysis Toolkit[2] (GATK v3.1). Bam file level recalibration was also performed using the BaseRecalibrator module from GATK.

Variant discovery and genotype calling of multiallelic substitutions, insertions and deletions were performed on all individuals globally using the UnifiedGenotyper and HaplotypeCaller modules from Genome Analysis Toolkit (GATK v3.1) as well as the FreeBayes variant caller (v9.9.2). The Ensemble variant calling pipeline (v0.2.2; http://bcb.io/2013/02/06/an-automated-ensemble-method-for-combining-and-evaluating-genomic-variants-from-multiple-callers/) was then implemented to integrate analysis results from the above three callers. Then, the Ensemble variant calling pipeline applies a machine learning algorithm called support vector machine to identify an optimal decision boundary on the basis of the variant calling results out of multiple variant callers, with an aim to improve the caller's receiver operating characteristic—in other words, a more balanced decision between false positives and true positives.

In addition, insertions and deletions were left-aligned at both post-alignment (BAM) and postvariant calling (VCF) levels using GATK's LeftAlignIndels and LeftAlignVariants modules, respectively.

We imported Northern and Western European ancestry (CEU), African ancestry (YRI), and Asian ancestry (ASI) (Han Chinese and Japanese) patients from 1000 Genomes (1 KG) for the admixture analysis. For each of the ethnic groups, we imported 419,035 SNPs overlapping to the rhabdo-myosarcoma (RMS) secondary cohort and filtered out monomorphic SNPs in any of the ethnic groups and ended up with 35,502 SNPs. We used these SNPs to estimate the ethnic components of the RMS secondary cohort. The estimation was made with an in-house imputation of the algorithm described in the work of Pritchard et al.[3] We considered CEU > 0.8 as White, YRI > 0.8 as Black or African American, ASI > 0.8 as Asian, and the rest as admix.

Genome Sequencing

Five hundred nanogram to 1 μg of genomic DNA was submitted to The Centre for Applied Genomics for genomic library preparation and genome sequencing. DNA samples were quantified using the Qubit High Sensitivity Assay, and purity was assessed using the Nanodrop OD260/280 ratio. Approximately 500-700 ng of DNA was used as input material for library preparation using the Illumina TruSeq PCR-free DNA Library Prep Kit following the manufacturer's recommended protocol. In brief, DNA was fragmented to 400 bp on average using sonication on a Covaris LE220 instrument. Fragmented DNA was end-repaired; A-tailed and indexed TruSeq Illumina adapters with overhang-T will be added to the DNA. Libraries were then validated on a Bioanalyzer DNA High Sensitivity chip to check for size and absence of primer dimers and quantified by quantitative PCR (qPCR) using Kapa Library Quantification Illumina/ABI Prism Kit protocol (KAPA Biosystems). Validated libraries were pooled in equimolar quantities and paired-end sequenced on an Illumina HiSeq X platform following Illumina's recommended protocol to generate paired-end reads of 150-bases in length. SNV and indel detection was performed using HaplotypeCaller in a gVCF model of GATK v.4.0.2.1.[4] All detected variants were filtered to remove false positive using the following

requirements: read position 10-90, strandedness 1%-99%, distance to 3′ > 20, homopolymer < 5, map quality difference < 30, read length difference < 25, and mismatch quality sum (MMQS) difference < 100. Variants with less than five alternate reads detected using bam-readcount were removed.

Automated and Manual Review of Variation

Any variants that had a read depth < 10 or allelic balance of heterozygous calls < 0.2 or > 0.8 were excluded from the analysis. Only variants with refGene nonsynonymous exonic and splicing were included in the analysis. All rare variants (minor allele frequency [MAF] < 1%) from the candidate gene list were classified into pathogenic (P), likely pathogenic (LP), variant of uncertain significance, likely benign, and benign first from ClinVar calls, made based on clinical laboratories meeting minimum requirements for data sharing to support quality assurance[5] (badged lab) by ClinGen (https://www.clinicalgenome.org/lablist/) using an archive database downloaded on May 20, 2018 (nonbadged lab calls were disregarded). At least one call from ClinVar badged lab was needed to classify variants and a majority rule applied. If no ClinVar calls were available, Human Gene Mutation Database (HGMD) (2019.1)[6] disease-causing mutation variants were subjected to manual review according to American College of Medical Genetics (ACMG)-Association of Molecular Pathology (AMP) guidelines.[7] Next, variation was interpreted by InterVar.[8] Final calls were prioritized in the following order: ClinVar, HGMD manual review, and InterVar (Appendix Fig 1). All the P/LP variants in BAM files were reviewed.

Multigene Panel Sequencing Analysis

Next-generation sequencing library preparation and custom-targeted capture.
Blood-derived genomic DNAs (100 ng) from 276 intermediate-risk RMS germline samples were sheared using a Covaris S2 instrument (Covaris, Woburn, MA). Libraries were prepared using the Ovation Ultralow Library System (NUGEN #0329) according to the manufacturer's instruction. Library enrichment for a 59-gene custom panel was done with the Roche SeqCap EZ Choice Library (cat #06266339001) and the SeqCap EZ Reagent Kit Plus v2 (NimbleGen #06-953-247-001) using the manufacturer's protocol. Individual libraries were combined into pools of 12 before hybridization. Captured libraries were sequenced on the Illumina HiSeq2000 channel using the HiSeq 101 Cycle Paired-End sequencing protocol. The 59-gene custom capture was selected based on their known contribution to hereditary cancer syndromes (59 genes include *ALK, APC, ATM, AXIN2, BAP1, BARD1, BMPR1A, BRCA1, BRCA2, BRIP1, CDH1, CDK4, CDKN2A, CHEK2, CTNNB1, DICER1, FLCN, MAX, MEN1, MLH1, MLH3, MRE11A, MSH2, MSH6, MUTYH, NBN, NF1, PALB2, PHOX2B, PKHD1, PMS1, PMS2, POLD1, POLE, POT1, PRSS1, PTCH1, PTEN, RAD50, RAD51B, RAD51C, RET, RINT1, SDHA, SDHAF2, SDHB, SDHC, SDHD, SMARCB1, STK11, SUFU, TCF7L2, TGFBR2, TMEM127, TP53, VHL, WT1, XRCC2, XRCC3*). Generated reads were aligned to the hg19 reference genome with Burrows-Wheeler Alignment Tool.[9] Base quality score recalibration, indel realignment, duplicate removal, and SNP/INDEL calling and other processing to variant calls were completed via the GATK.[10] Samples that failed library preparation were excluded from the analysis (n = 18) as well as samples with less than an average of 10× coverage over the panel (n = 7) as determined by bedtools.[11,17]

Structural Variant Analysis

Filtering and annotation of deletions, duplications, and inversions.
An ensemble approach was taken, retaining only the deletions, duplications, and inversions called by both DELLY and Manta, based on the following size-specific constraints for breakpoint consensus between the two tools: breakpoints within 100 bp for structural variation (SV) ≤ 10 kbp, breakpoints within 1 kbp for SVs ≥ 10 kbp and ≤ 50 kbp, and breakpoints within 10 kbp for SVs ≥ 50 kbp. A high-quality SV set was obtained by applying additional filtering criteria. A panel of normal controls was created to remove technical false positives using 72 germline genomes sequenced on the same HiSeqX platform (150bp paired end sequencing, minimum of 30× depth coverage). SVs present in ≥ 3 panel of normal genomes were removed. Annotation was performed with AnnotSV and subsequently filtered to remove SVs with an AnnotSV score < 3 and present in publicly available normal databases: 1000 Genomes Project and gnomAD. The remaining exonic SVs were subject to manual review in Integrative Genomics Viewer.

PCR validation of structural variation.
In all instances, amplification was performed in a 50-μL reaction volume, and PCR products were examined by agarose gel electrophoresis. The reaction consisted of 25 μL of a standard Taq master mix, 2.5 μL each of both the forward and reverse primers (10 μM), and 100 ng of the relevant template DNA (patient DNA and control human genomic DNA), with the remaining volume being nuclease-free water. Primer sequences for *ATM* are 5′- TGGGTAATAATTTCCAAGTGAAACTCCCCA-3′ and 5′-GAGTTGCAGGGGGATAATAGTGATGATGTG-3′, and CDKN1C are 5′-TCAGCTTTGTTTACGTCGCCGCGCAATGTG-3′ and 5′-AGGCGCCGG-AGCAGCTGCCTAGT-3′. An initial denaturation of 30 seconds was performed at 98°C followed by 35 cycles of amplification, which consisted of a 10-second denaturation step performed at 98°C, a 10-second annealing step performed at 60°C, and an extension step of 30 seconds or 3 minutes, for *CDKN1C* and *ATM*, respectively (Appendix Figs 2A and A3), performed at 72°C. A final extension of 5 minutes was performed at 72°C.

Statistical Analysis

Survival/multivariate analysis.
Event-free survival (EFS) was defined as time from enrollment on a Children's Oncology Group study until event (relapse, SMN, or death) or until last contact. Patients who did not experience an EFS event were considered censored at the time of last contact. Overall survival was defined as time from enrollment on a COG study until death or until last contact. Patients who were alive at last contact were considered censored at the time of last contact.

Each multivariate model was produced using stepwise selection, starting with a model containing only the predictor variable being tested (which was forced to be in the model) and allowing for the addition of the following demographic and clinical variables: age at enrollment (as a categorical variable), sex, race, ethnicity, RMS type, RMS stage, N-stage, and T-stage. The final model included the predictor variable being tested (which was forced to be in the model) and any demographic variables found to be significant. The P-value presented for each of the multivariate models is the type 3 P value from the Cox proportional hazard regression for the predictor variable given that any significant demographic and diagnostic variables are also in the model. Each predictor variable was a 0/1 categorical variable with 1 indicating the subject being a P/LP carrier in 130 CSGs or 102 AD, AD/AR genes, and 0 indicating the subject not being a P/LP carrier.

Age at enrollment was defined as fractional years from birth until enrollment on a COG trial. It was treated as a continuous variable. Age category at enrollment was defined as years from birth until enrollment on a COG trial. It was treated as a categorical variable with the following categories: < 1 year old, 1-9 years old, and ≥ 10 years old. Patient sex was treated as a categorical variable with two categories (male and female). Patient race was treated as a categorical variable with four categories (White, Black, Others, and Unknown). Patient ethnicity was treated as a categorical variable with three categories (Hispanic, Non-Hispanic, and Unknown). RMS type was treated as a categorical variable with three categories (Alveolar, Embryonal, and Not Otherwise Specified). RMS stage was treated as a categorical variable with three categories (stage I, stage II, and stage III). N-stage (nodal involvement) was treated as a categorical variable with 3 categories (No/N_0, Yes/N_1, and Not Evaluated/Unknown). T-stage was treated as a categorical variable with 2 categories (T_1 and T_2).

Multigene panel comparison.
Germline results were compared with the results of a subset of patients from the same COG discovery cohort, which were run earlier on a custom-made multigene panel from the University of Utah (see Supplemental Methods).

REFERENCES

1. Lohse M, Bolger AM, Nagel A, et al: RobiNA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res40:W622-W627, 2012
2. DePristo MA, Banks E, Poplin R, et al: A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491-498, 2011
3. Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. Genetics 155:945-959, 2000
4. McKenna A, Hanna M, Banks E, et al: The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-1303, 2010
5. Landrum MJ, Lee JM, Benson M, et al: ClinVar: Improving access to variant interpretations and supporting evidence. Nucleic Acids Res 46:D1062-D1067, 2018
6. Stenson PD, Ball EV, Mort M, et al: Human gene mutation database (HGMD®): 2003 update. Hum Mutat 21:577-581, 2003
7. Richards S, Aziz N, Bale S, et al: Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405-424, 2015
8. Li Q, Wang K: InterVar: Clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. Am J Hum Genet 100:267-280, 2017
9. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760, 2009
10. DePristo MA, Banks E, Poplin R, et al: A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491-498, 2011
11. Quinlan AR, Hall IM: BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842, 2010
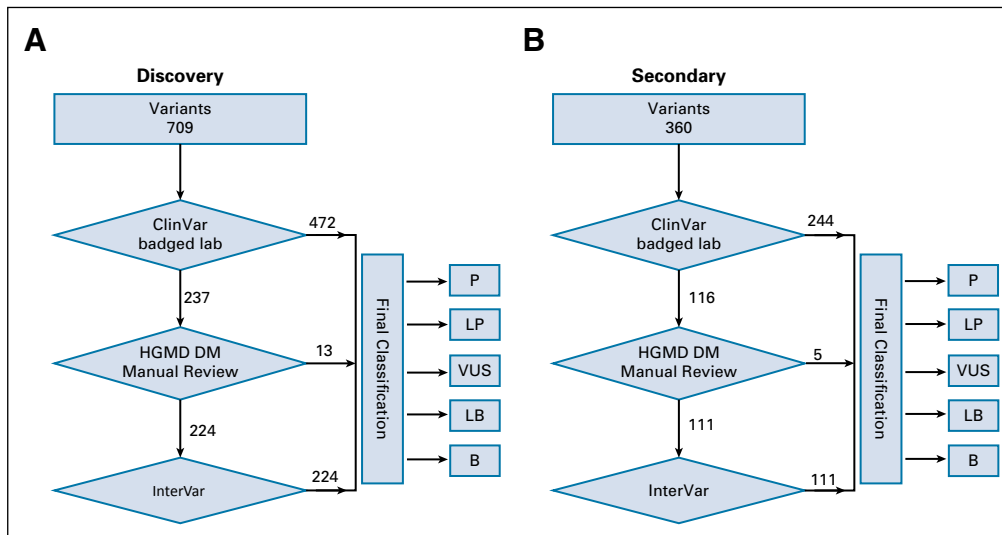
FIG A1. Flow diagram of unique variant classification. (A) Discovery cohort, (B) secondary cohort. B, benign; DM, disease causing; HGMD, human gene mutation database; LB, likely benign; LP, likely pathogenic; P, pathogenic; VUS, variant of unknown significance.
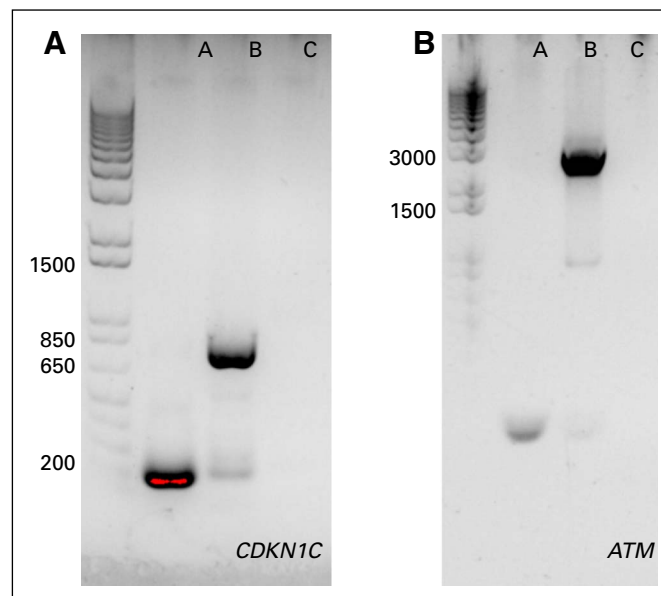

FIG A2. Structural variant PCR validation using 1% agarose gel in tris-acetate-EDTA buffer: (A) *CDKN1C* and (B) *ATM*. Columns: A, patient DNA; B, control genomic DNA; C, no template.
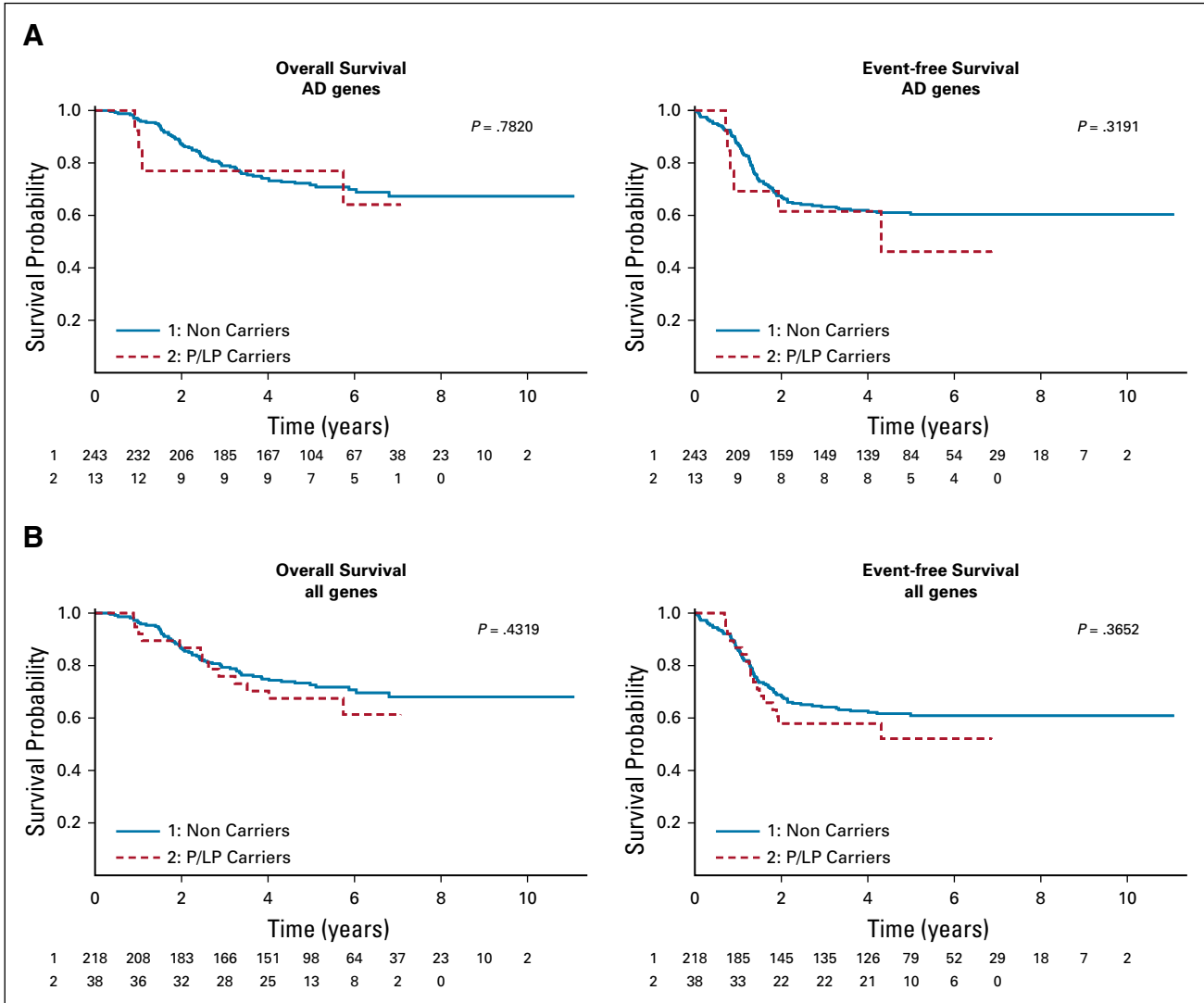
FIG A3. (A) Overall survival analysis and event-free survival analysis in AD genes only and (B) all genes in Discovery cohort.