OXFORD

Germline genetic variants and pediatric rhabdomyosarcoma outcomes: a report from the Children's Oncology Group

Bailey A. Martin-Giacalone (iD), PhD,[1,2] Melissa A. Richard, PhD,[2] Michael E. Scheurer, PhD,[2,3] Javed Khan, PhD,[4] Pagna Sok, MPH,[2] Priya B. Shetty, PhD,[2] Stephen J. Chanock, PhD,[5] Shengchao Alfred Li, MS,[6] Meredith Yeager, PhD,[6] Deborah A. Marquez-Do, BS,[2] Donald A. Barkauskas (iD), PhD,[7,8] David Hall (iD), MS,[8] Matthew T. McEvoy, MD,[2] Austin L. Brown, PhD,[2] Aniko Sabo (iD), PhD,[9] Paul Scheet, PhD,[10] Chad D. Huff, PhD,[10] Stephen X. Skapek, MD,[11] Douglas S. Hawkins (iD), MD,[12] Rajkumar Venkatramani, MD,[2] Lisa Mirabello, PhD,[13] Philip J. Lupo (iD), PhD[2,*]

[1]Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA
[2]Section of Hematology-Oncology, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA
[3]Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA
[4]Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
[5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
[6]Frederick National Laboratory for Cancer Research, Frederick, MD, USA
[7]Department of Population and Public Health Sciences, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA
[8]QuadW Childhood Sarcoma Biostatistics and Annotation Office, Children's Oncology Group, Monrovia, CA, USA
[9]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA
[10]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
[11]Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX, USA
[12]Division of Hematology-Oncology, Department of Pediatrics, Seattle Children's Hospital, University of Washington, Seattle, WA, USA
[13]Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MA, USA

*Correspondence to: Philip J. Lupo, PhD, Department of Pediatrics, Baylor College of Medicine, One Baylor Plaza, MS: BCM622, Houston, TX 77030, USA (e-mail: philip.lupo@bcm.edu).

Abstract

Background: Relative to other pediatric cancers, survival for rhabdomyosarcoma (RMS) has not improved in recent decades, suggesting the need to enhance risk stratification. Therefore, we conducted a genome-wide association study for event-free survival (EFS) and overall survival (OS) to identify genetic variants associated with outcomes in individuals with RMS.

Methods: The study included 920 individuals with newly diagnosed RMS who were enrolled in Children's Oncology Group protocols. To assess the association of each single nucleotide polymorphism (SNP) with EFS and OS, we estimated hazard ratios (HRs) and 95% confidence intervals (CIs) using multivariable Cox proportional hazards models, adjusted for clinical covariates. All statistical tests were two sided. We also performed stratified analyses by histological subtype (alveolar and embryonal RMS) and carried out sensitivity analyses of statistically significant SNPs by *PAX3/7-FOXO1* fusion status and genetic ancestry group.

Results: We identified that rs17321084 was associated with worse EFS (HR = 2.01, 95% CI = 1.59 to 2.53, $P = 5.39 \times 10^{-9}$) and rs10094840 was associated with worse OS (HR = 1.84, 95% CI = 1.48 to 2.27, $P = 2.13 \times 10^{-8}$). Using publicly available data, we found that rs17321084 lies in a binding region for transcription factors GATA2 and GATA3, and rs10094840 is associated with *SPAG1* and *RNF19A* expression. We also identified that *CTNNA3* rs2135732 (HR = 3.75, 95% CI = 2.34 to 5.99, $P = 3.54 \times 10^{-8}$) and *MED31* rs74504320 (HR = 3.21, 95% CI = 2.12 to 4.86, $P = 3.60 \times 10^{-8}$) were associated with worse OS among individuals with alveolar RMS.

Conclusions: We demonstrated that common germline variants are associated with EFS and OS among individuals with RMS. Additional replication and investigation of these SNP effects may further support their consideration in risk stratification protocols.

Rhabdomyosarcoma (RMS) is the most common soft-tissue sarcoma of childhood and is commonly classified into two major histological subtypes: embryonal RMS (ERMS) and alveolar RMS (ARMS). Importantly, 80% of ARMS have a chromosomal translocation that results in fusion of the gene *PAX3* or *PAX7* with *FOXO1*. Several studies have shown that children with PAX/FOXO1 fusion-positive RMS have statistically significantly worse survival than those with fusion-negative RMS (1-3). These findings led to the incorporation of fusion status in risk stratification for Children's Oncology Group (COG) protocols (4).

Clinical trials testing novel therapeutics for children with RMS have largely been unsuccessful in increasing survival outcomes over the past several decades; this is especially true for patients with intermediate- and high-risk RMS (5-7). One strategy to address poor outcomes among these children is to enhance risk-stratified diagnostic protocols through the incorporation of

prognostic genomic markers. For example, somatic mutations in *TP53* and *MYOD1* are associated with poor survival for children with fusion-negative RMS (8); these markers are being included in upcoming risk-stratified COG clinical trials to identify individuals with higher risk of poor outcomes (4).

Germline variants play a role in RMS predisposition (9); whether they affect RMS outcomes remains unclear. Addressing these questions could help design COG risk-stratified clinical trials that assess whether individuals with certain germline variants might benefit from more aggressive treatment. Further, providing insight into the etiology of response could inform novel therapeutic developments for RMS. Although genome-wide association studies (GWAS) have detected common variants associated with poor outcomes for other pediatric cancers (10-13), there have been no published GWAS of outcomes among individuals with RMS. Therefore, we conducted a GWAS for event-free survival (EFS) and overall survival (OS) in a large, unselected cohort of individuals with RMS.

Methods
Study population
The Institutional Review Board for Human Subjects Research at Baylor College of Medicine approved this study. The initial cohort comprised 924 individuals (age <40 years) with newly diagnosed RMS who were consented to the COG soft-tissue sarcoma biobanking protocol D9902. Additional information on data collection can be found in the Supplementary Methods (available online).

The analysis included all risk groups. We recoded case histology that was not ERMS or ARMS. Specifically, we coded botryoid (n = 74) and spindle cell (n = 49) RMS as ERMS because individuals with these subtypes have similar outcomes (14). Mixed RMS, RMS with ganglionic differentiation, and RMS not otherwise specified were coded as "other." In sensitivity analyses by PAX/FOXO1 fusion status, all ERMS were coded as fusion-negative RMS because virtually all ERMS lack the *PAX3/7-FOXO1* fusion gene (1,14).

Genotyping and quality control
All individuals underwent genome-wide genotyping on Illumina BeadChip arrays. Genotyping and quality control were performed as described in the Supplementary Methods (available online). After quality control, the total number of single nucleotide polymorphisms (SNPs) was 159 354, and the total number of individuals in the study cohort was 920. We imputed the dataset using the Michigan Imputation Server (15) with reference data from the Haplotype Reference Consortium. The number of SNPs with an imputation quality score (r^2) of at least 0.6 was 5 387 542; these SNPs comprised the final dataset.

Statistical analysis
The primary endpoints of the study were: 1) EFS, defined as the time from date of study enrollment to tumor recurrence or progression, secondary malignancy, or death due to any cause; and 2) OS, defined as the time from study enrollment to death due to any cause. Individuals without an event were censored at time of last contact.

Using the gwasurvivr 1.12.0 R package (16), we conducted Cox proportional hazards regression to calculate a hazard ratio (HR), 95% confidence interval (CI), and P value for each SNP with EFS and OS in an additive model. In the Cox regression model, age at diagnosis (categorical: <1 year, 1-9 years, ≥10 years), tumor stage (categorical: 1, 2, 3, 4), and histological subtype (categorical: ARMS, ERMS, other) were statistically significantly associated

(P < .05) with outcome across the cohort (Supplementary Table 1, available online). Therefore, the Cox models were adjusted for these factors. The top five genetically estimated principal components were generated using PLINK (17) and were also included in the final model to account for population stratification or differences in allele frequency due to the presence of underlying ancestral subgroups (18). We then generated quantile-quantile plots to assess residual genomic inflation (Supplementary Figure 1, available online). PAX/FOXO1 fusion status was not included in the final model because this variable was unknown for 83.4% of the study cohort.

We defined genome-wide, statistically significant associations as those with a P value less than 5×10^{-8}, which is a Bonferroni correction for multiple testing across 1 million independent segments of the genome (19). To identify independent SNP associations, we performed conditional analyses using GCTA software (20). For genome-wide statistically significant associations, we used the Kaplan-Meier estimator and log-rank test to evaluate differences in survival by genotype. Based on calculating and visualizing Schoenfeld residuals, we found no statistically significant violations of the proportional hazards assumption. All statistical tests were two sided and conducted in R version 4.0.4.

Analytic groups
We conducted a GWAS across all subtypes and then stratified the GWAS by the two major histological subtypes: ARMS and ERMS. We also performed post hoc analyses using logistic regression models to determine whether the frequency of genome-wide statistically significant SNPs across the entire cohort differed by histological subtype.

Sensitivity analyses
We conducted sensitivity analyses of SNP associations that were genome-wide statistically significant across the entire cohort. SNPs were evaluated within Admixed American, African, and European genetic ancestry groups, which were defined as described in the Supplementary Methods (available online). In a separate analysis, we evaluated SNP effects by PAX/FOXO1 fusion status.

Functional annotation
We visualized linkage disequilibrium patterns and generated regional association plots using LocusZoom v0.14.0 (21). Additionally, we utilized publicly available databases (described in Supplementary Methods, available online) to provide genomic annotation for each locus that contained a genome-wide statistically significant SNP. We also queried the Genotype-Tissue Expression (GTEx) Portal (22) for expression quantitative trait loci (eQTL) and splicing QTL (sQTL) across the 49 tissues analyzed in the v8 release.

Results
Study population
The characteristics of the 920 individuals in the final study cohort (Table 1) are consistent with previous population-based studies of RMS (23,24). The median age at diagnosis was 5.6 years (range = 3.7 days to 37.5 years), and there was a male predominance (1.4 males to 1 female). The most common histological subtype was ERMS followed by ARMS. Of the individuals with ARMS who had known PAX/FOXO1 fusion status, 78.6% had fusion-positive ARMS. The median follow-up time was 4.7 years (range = 1.1 days to 14.9 years).

Table 1. Demographic and clinical characteristics of the rhabdomyosarcoma study cohort

Characteristic	No. (%)
Histological subtype	
Alveolar	268 (29.1)
Embryonal	544 (59.1)
Other	108 (11.8)
Sex	
Female	384 (41.7)
Male	536 (58.3)
Race and ethnicity	
Hispanic or Latino	135 (14.7)
Not Hispanic or Latino	
American Indian or Alaska Native	8 (0.9)
Asian	32 (3.5)
Black or African American	110 (12.0)
Native Hawaiian or Pacific Islander	3 (0.3)
White	557 (60.5)
Unknown	75 (8.1)
Age at diagnosis, y	
<1	54 (5.9)
1-9	556 (60.4)
≥10	310 (33.7)
Tumor stage	
1	271 (29.5)
2	160 (17.4)
3	311 (33.8)
4	178 (19.3)
Event-free survival status	
Event	359 (39.0)
No event	561 (61.0)
Overall survival status	
Event	273 (29.7)
No event	647 (70.3)

SNPs associated with EFS

We identified a genome-wide statistically significant association between rs17321084 (chr8q24.13) and EFS (Figure 1, A); rs17321084 was not in strong linkage ($r^2 < 0.8$; 1000 Genomes Phase 1 European population) with other SNPs in the region [Figure 2, A; HaploReg (25)]. The T allele was statistically significantly associated with worse EFS (HR = 2.01, 95% CI = 1.59 to 2.53, $P = 5.39 \times 10^{-9}$, Table 2). Further, individuals who were heterozygous or homozygous for the T allele had statistically significantly worse EFS (P < .0001) compared with those who were homozygous for the C allele (Figure 2, B). The odds of having the T allele were greater in those with ARMS compared with those with ERMS (odds ratio = 1.07, 95% CI = 1.01 to 1.13, P = .01). In sensitivity analysis by histology, the effect estimates were consistent with our initial findings (Supplementary Table 2, available online).

There were no SNP-gene expression associations reported in GTEx for rs17321084. However, rs17321084 lies in a binding region for transcription factors GATA2 and GATA3 [neuroblastoma cell line; ENCODE GEO: GSM935589 (26)]. Additionally, publicly available RMS tissue microarray and OS data on individuals with RMS [Oncogenomics DB (27)] showed that those with high tumor expression of GATA2 or GATA3 had worse OS compared with those with low expression (GATA2: $P = 3.63 \times 10^{-4}$, GATA3: $P = 5.92 \times 10^{-4}$).

At a lower statistical significance threshold ($P < 1 \times 10^{-6}$), the SNP with the strongest association with EFS was rs113830923 (ch12p12.1; Figure 2, C). The G allele was statistically significantly associated with worse EFS (HR = 1.97, 95% CI = 1.53 to 2.53, $P = 1.23 \times 10^{-7}$; Supplementary Table 3, available online).
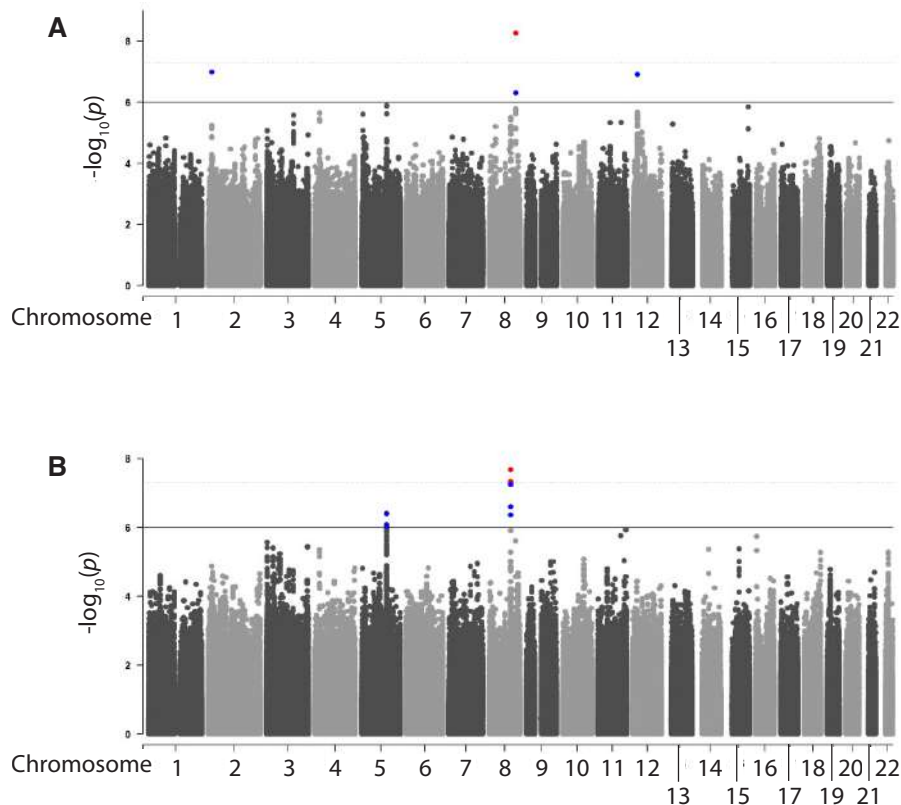

Figure 1. Association of common variants with survival outcomes among 920 individuals with rhabdomyosarcoma. Manhattan plots of a genome-wide association study of A) event-free survival and B) overall survival, displaying genome-wide statistically significant single nucleotide polymorphism associations on chromosome 8.
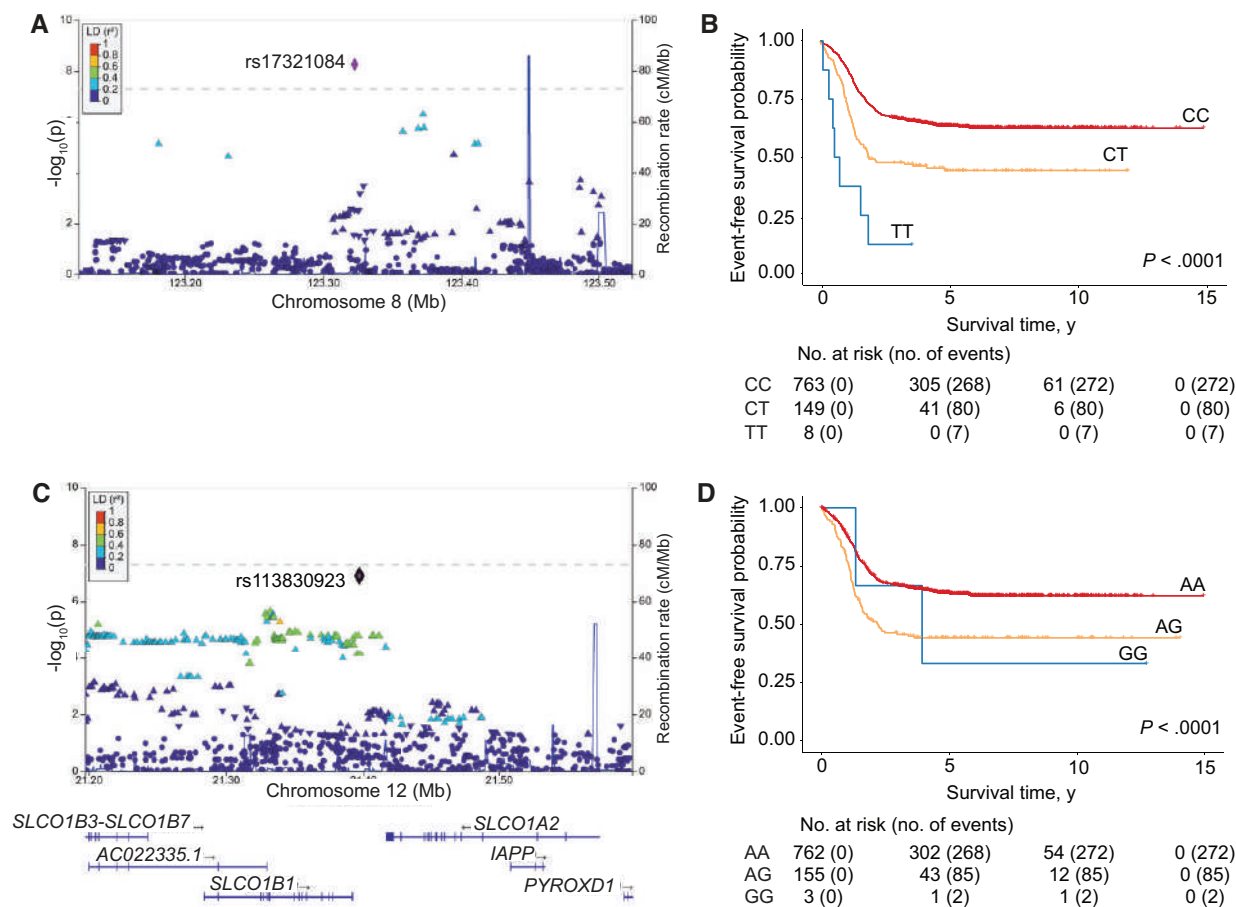
Figure 2. Association of single nucleotide variants with event-free survival (EFS) of 920 individuals with rhabdomyosarcoma. Regional association plot displaying linkage disequilibrium (LD) and recombination hotspots for the: A) chr8q24.13 locus, which harbors rs17321084; and B) Kaplan-Meier curve of EFS by genotype (CC, CT, TT) of rs17321084. Regional association plot for the C) chr12p12.1 locus, which harbors rs113830923, and D) Kaplan-Meier curve of EFS by genotype (AA, AG, GG) of rs113830923. Mb = Megabase.

Table 2. Association of genome-wide statistically significant single nucleotide polymorphisms (SNPs) with event-free survival (EFS) and overall survival (OS) among individuals with rhabdomyosarcoma[a]

Outcome	Subtype	Chr	Position	SNP	Alleles	MAF	HR (95% CI)	P
EFS	All	8	123323659	rs17321084	C/T	8.7%	2.01 (1.59 to 2.53)	5.39×10^{-9}
OS	All	8	101374772	rs10094840	G/A	16.7%	1.84 (1.48 to 2.27)	2.13×10^{-8}
	ARMS	10	68084977	rs2135732	G/A	10.0%	3.75 (2.34 to 5.99)	3.54×10^{-8}
	ARMS	17	6554921	rs74504320	A/G	8.9%	3.21 (2.12 to 4.86)	3.60×10^{-8}

[a] ARMS = alveolar rhabdomyosarcoma; Chr = chromosome; CI = confidence interval; HR = hazard ratio; MAF = sample minor allele frequency. Position is based on reference hg19. Alleles are shown as major/minor.

Individuals who were heterozygous or homozygous for the G allele had statistically significantly worse EFS ($P < .0001$) compared with those who were homozygous for the A allele (Figure 2, D). There were no statistically significant differences in the frequency of the risk allele by histological subtype.

SNP rs113830923 lies approximately 5 kb from *SLCO1B1*, a gene that encodes the organic anion transporter polypeptide 1B1 (OATP1B1) (28). This SNP is an eQTL for *SLCO1B1* and *SLCO1B7* in tibial nerve tissue; increasing copies of the G allele were associated with higher *SLCO1B1* expression (GTEx). However, there were no statistically significant eQTLs identified in skeletal muscle or other tissues derived from those of which RMS is thought to arise. Data from Oncogenomics DB show that individuals with RMS who have low somatic expression of *SLCO1B1* have worse OS than those with high expression ($P = 8.52 \times 10^{-3}$).

SNPs associated with OS

We identified a genome-wide statistically significant association between rs10094840 (chr8q22.2) and OS (Figure 1, B). Based on conditional analysis, rs10094840 was independent of other SNPs in the region (Figure 3, A). The A allele was statistically significantly associated with worse OS (HR = 1.84, 95% CI = 1.48 to 2.27, $P = 2.13 \times 10^{-8}$; Table 2). Individuals who were heterozygous or homozygous for the A allele had statistically significantly worse OS compared with those who were homozygous for the G allele ($P = .0001$; Figure 3, B). There was no difference in the frequency of the A allele by histological subtype.

The rs10094840 A allele is associated with higher *SPAG1* (a sperm-associated ciliary protein) expression in blood and mammary tissues and lower *RNF19A* (an E3 ubiquitin ligase) expression in the cerebellum (GTEx). We also found that this SNP lies
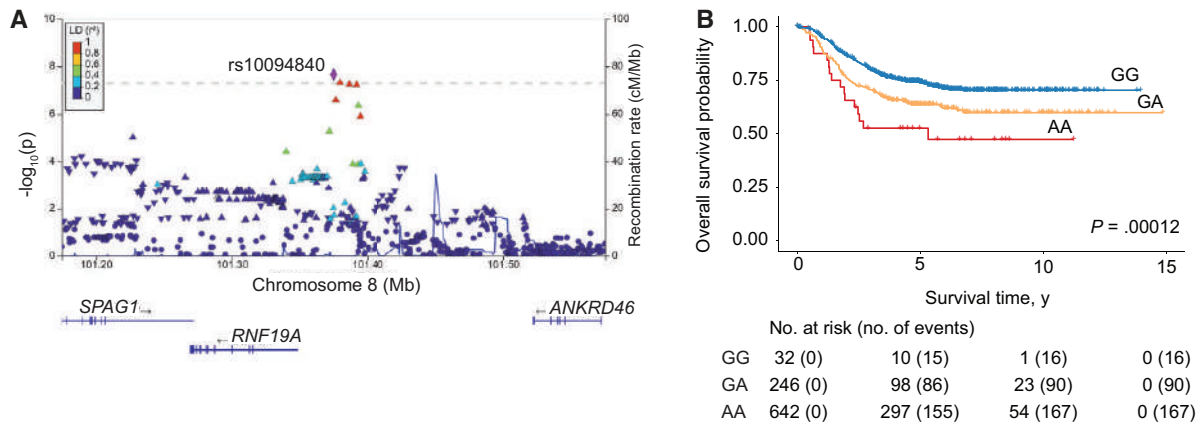
Figure 3. Association of single nucleotide variants with overall survival (OS) of 920 individuals with rhabdomyosarcoma. Regional association plot for the A) chr8q22.2 locus, which harbors rs10094840, and B) Kaplan-Meier curve of OS by genotype (GG, GA, AA) of rs10094840. Mb = Megabase.
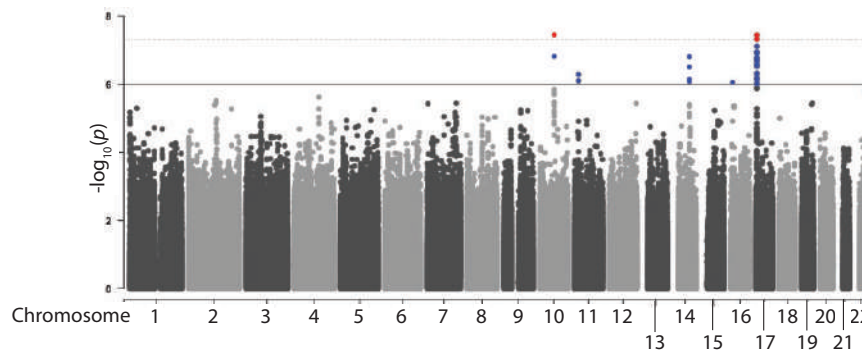

Figure 4. Association of common variants with overall survival (OS) among 268 individuals with alveolar rhabdomyosarcoma. Manhattan plot of a genome-wide association study of OS displaying genome-wide statistically significant single nucleotide polymorphism associations on chromosomes 10 and 17.

within a DNase I hypersensitivity site in embryonic skeletal muscle tissues of the leg (Roadmap Epigenomics GEO: GSM1027333). Oncogenomics DB data show that high RMS expression of *SPAG1* is associated with worse survival (P = .01), and low somatic expression of *RNF19A* is associated with worse survival (P = 4.33×10^{-3}).

SNPs associated with outcomes by histological subtype

We identified two genome-wide statistically significant associations with OS in individuals with ARMS (Figure 4). In individuals with ERMS, there were no genome-wide statistically significant SNP associations with either outcome. Estimates for SNPs that met a lower statistical significance threshold (P $< 1 \times 10^{-6}$) in ERMS or ARMS cohorts are provided in Supplementary Tables 4 and 5 (available online).

SNP rs2135732 (ch10q21.3) was statistically significantly associated with OS among individuals with ARMS (Figure 5, A). The A allele was statistically significantly associated with worse OS (HR = 3.75, 95% CI = 2.34 to 5.99, P = 3.54×10^{-8}; Table 2). Individuals with ARMS who were heterozygous or homozygous for the A allele had statistically significantly worse OS compared with those who were homozygous for the G allele (P = .002; Figure 5, B). This SNP lies in an intron of *CTNNA3* and is also located in a DNase I hypersensitivity site (embryonic skeletal muscle tissues of the arm; Roadmap Epigenomics GEO: GSM1027349, GSM774223), although there were no eQTLs or sQTLs reported for this SNP (GTEx).

We also identified a statistically significant association between rs74504320 (ch17p13.1) and OS among individuals with ARMS (Figure 5, C). SNP rs74504320 was independent of other SNPs in the region. The G allele was associated with worse OS (HR = 3.21, 95% CI = 2.12 to 4.86, P = 3.60×10^{-8}; Table 2). Among individuals with ARMS, those who were heterozygous or homozygous for the G allele had statistically significantly worse OS (P < .0001) compared with those who were homozygous for the A allele (Figure 5, D).

This SNP lies within the 5' untranslated region of transcriptional regulator *MED31* and is an eQTL and sQTL for several genes in the region, including *MED31*, across numerous tissues (GTEx, Supplementary Table 6, available online). This SNP also lies within a DNase I hypersensitivity site across tissues, including those of embryonic skeletal muscle (Roadmap Epigenomics GEO: GSM878618, GSM1059533); this site is a dense binding region for more than 50 transcription factors, including POLR2A, a major subunit of RNA polymerase II (ENCODE). Based on Oncogenomics DB data, individuals with RMS who have high somatic expression of *MED31* have worse survival than those with low expression (P = .01).

Sensitivity analyses

We carried out a sensitivity analysis of the genome-wide statistically significant SNPs in cohorts of individuals with *PAX3/7-FOXO1* fusion-negative (n = 576) and fusion-positive (n = 115) RMS. None of the SNPs were statistically significant in the sensitivity analysis, although effect estimates were consistent (Supplementary Table 2, available online).
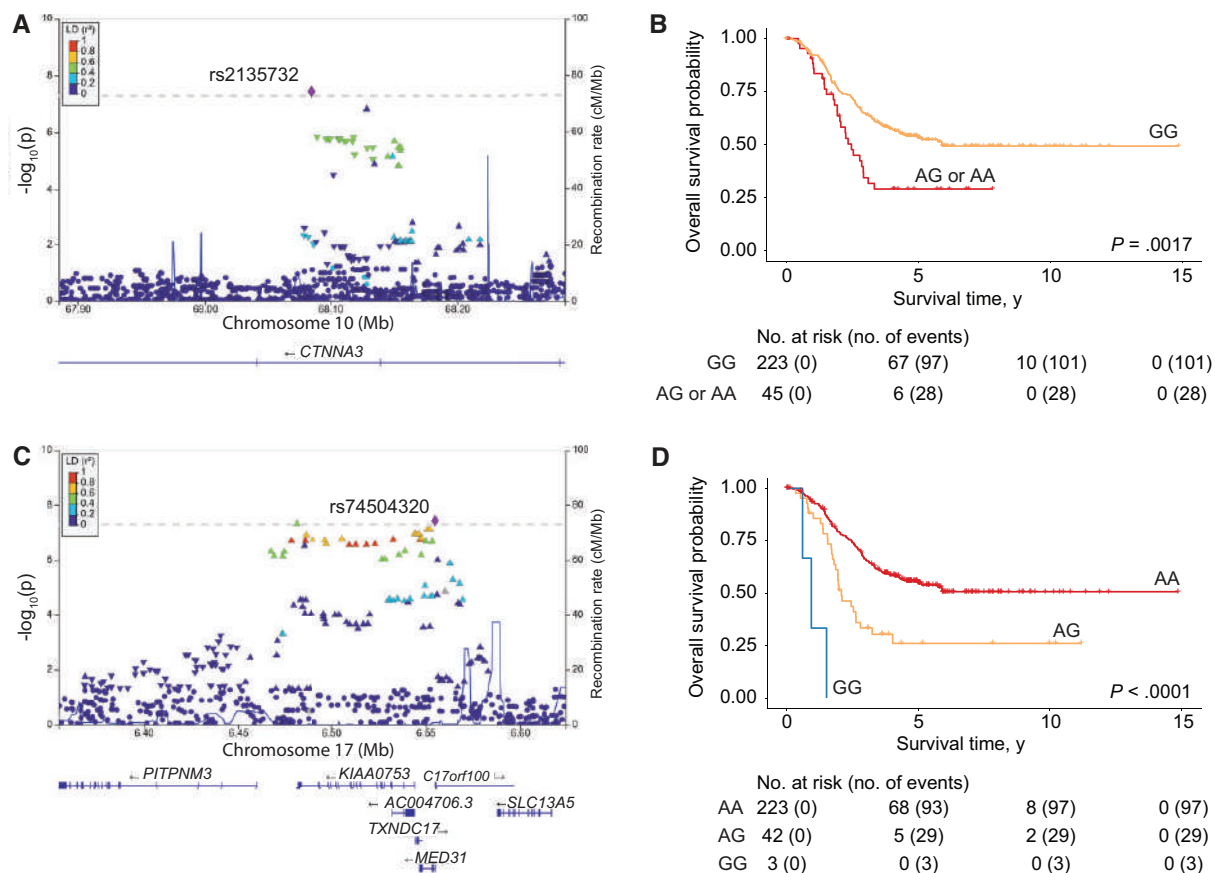
Figure 5. Association of single nucleotide variants with overall survival (OS) of 268 individuals with alveolar rhabdomyosarcoma. Regional association plot displaying linkage disequilibrium (LD) and recombination hotspots for the A) chr10q21.3 locus, which harbors rs2135732; and B) Kaplan-Meier curve of OS by genotype (GG, AG, or AA) of rs2135732. Individuals with AG or AA genotype were grouped because there was only 1 individual with AA genotype. Regional association plot for the C) chr17p13.1 locus, which harbors rs74504320, and D) Kaplan-Meier curve of OS by genotype (AA, AG, GG) of rs74504320.

Table 3. Effect estimates of genome-wide statistically significant single nucleotide polymorphisms (SNPs) with event-free survival (EFS) and overall survival (OS) by genetic ancestry[a]

	Admixed American ancestry (N = 210)			African ancestry (N = 96)			European ancestry (N = 534)		
Outcome/SNP	MAF	HR (95% CI)	P	MAF	HR (95% CI)	P	MAF	HR (95% CI)	P
EFS: rs17321084	7.0%	1.56 (0.83 to 2.92)	.16	1.3%	4.32 (0.69 to 27.02)	.12	11.0%	1.98 (1.52 to 2.59)	5.80×10^{-7}
OS: rs10094840	13.8%	2.09 (1.38 to 3.16)	5.29×10^{-4}	14.7%	2.15 (0.91 to 5.07)	.08	18.9%	1.81 (1.37 to 2.37)	2.25×10^{-5}

[a] CI = confidence interval; HR = hazard ratio; MAF = sample minor allele frequency.

We also analyzed the effects of genome-wide statistically significant SNPs in three major genetic ancestry groups. The effects of the SNPs within each genetic ancestry group were similar in magnitude and direction to the effects across the entire cohort (Table 3). Though not statistically significant, we did observe that in the African ancestry group, the rs17321084 risk allele was of lower frequency and the effect estimate was twofold greater than the estimate across the entire cohort and other ancestry groups (minor allele frequency = 1.3%; HR = 4.32, 95% CI = 0.69 to 27.02, P = .12).

Discussion

In our GWAS of survival outcomes among individuals with RMS, we identified genome-wide statistically significant associations for EFS and OS. Supported by functional data, we also identified a locus on chromosome 12 that suggests association with EFS. In histology-specific analyses, we identified genome-wide statistically significant loci on chromosomes 10 and 17 that were associated with OS in individuals with ARMS.

The SNP rs17321084 lies in a binding region for transcription factors GATA2 and GATA3. GWAS of acute lymphoblastic leukemia have identified SNPs in GATA3 that are associated with poor outcomes such as minimal residual disease and relapse (12,13). Because GATA2 and GATA3 expression is associated with survival for individuals with RMS [Oncogenomics DB (27)], future efforts could explore the mechanism by which this germline SNP might result in differential GATA2 or GATA3 expression or regulation. We also found that the frequency of this variant was statistically significantly higher in individuals with ARMS compared with individuals with ERMS. To determine whether this observation is

driving the association with EFS, additional validation and analysis in independent cohorts is needed.

We also identified a SNP (rs113830923) near *SLCO1B1* that was associated with worse EFS ($P < 1 \times 10^{-6}$) among individuals with RMS. *SLCO1B1* encodes a hepatic transporter (OATP1B1) for a wide range of compounds, including chemotherapeutics (28). Pharmacogenetic GWAS of children with acute lymphoblastic leukemia have identified SNPs in *SLCO1B1* that are associated with methotrexate-related toxicity (29-31). In the RMS literature, Sakaguchi et al. (32) reported an individual with ARMS who harbored a haplotype containing two missense SNPs in *SLCO1B1* in addition to a SNP in *UGT1A1*. The authors hypothesized that because OATP1B1 specifically transports SN-38, an active metabolite of irinotecan, these germline variants contributed to the individual experiencing irinotecan-related severe neutropenia while being treated on a COG protocol. Interestingly, *Xenopus* models with the individual's *SLCO1B1* haplotype have displayed reduced OATP1B1 activity and uptake of SN-38 compared with wild-type models (33).

Because irinotecan is a standard component of chemotherapy for intermediate-risk RMS (4), we evaluated the effect of rs113830923 on EFS in the ARST0531 intermediate-risk cohort. The magnitude and direction of the effect were consistent in the subset of patients treated with irinotecan (Supplementary Table 7, available online). We did not have consistently collected data on toxicities in the study cohort. Given our findings, future work could assess whether this SNP association is driven by specific treatment-related toxicities that are associated with outcome. Further exploration could also determine how rs113830923 might influence RMS *SLCO1B1* expression.

We found that rs10094840 was associated with expression of *SPAG1* and *RNF19A* in certain tissues analyzed by GTEx. Some sperm-associated antigen genes have been associated with poor osteosarcoma survival (34), while the oncogenic properties of both *SPAG1* and *RNF19A* are emerging in other cancers (35,36). Because rs10094840 lies within a DNase I hypersensitivity site in embryonic skeletal muscle tissues and somatic expression of these genes is associated with RMS survival [Oncogenomics DB (27)], additional work could consider further characterizing the mechanism underlying this germline association.

CTNNA3 rs2135732 was strongly associated with OS in individuals with ARMS. *CTNNA3*, an α-catenin, is thought to act as a tumor suppressor in carcinomas (37). Although *CTNNA3* has not been linked to RMS, mutations in *CTNNB1* are frequent in ERMS (38). Because α-catenin has the ability to bind β-catenin (encoded by *CTNNB1*) to stabilize cell adhesion junctions (39), further work could include characterizing the relationship between α- and β-catenin in RMS development and outcomes.

We also found a strong association between rs74504320 and OS in individuals with ARMS. This SNP lies within the 5' untranslated region of *MED31*, a subunit of the Mediator complex, which regulates transcription by binding RNA polymerase II (40). Thus, this SNP lies in a binding site for more than 50 transcription factors, including POLR2A, the largest subunit of RNA polymerase II. *MED31* overexpression is involved in osteosarcoma cell proliferation (41,42), and almost all Mediator subunits have been implicated in cancer (43). Future work could evaluate the specific effects of germline SNPs in *MED31* on transcriptional regulation in ARMS cell lines.

We carried out sensitivity analyses by fusion status and across three major genetic ancestry groups. The effects were largely consistent with our initial analyses in magnitude and direction. We did note that for rs17321084, the effect was stronger in individuals with African ancestry. Although our sample size was small, this result suggests that rs17321084 may play a role in RMS outcomes in this population. Future replication studies could further explore this finding and consider evaluating the effects of ancestry-specific variants on outcome.

This study is not without limitations. We were unable to control for the effects of treatment on survival outcomes because we were unable to obtain these data for 80.3% of individuals in the study. Systemic treatment across risk groups, however, relies on a common chemotherapeutic background, which has largely remained unchanged since the 1970s (4,44). Additionally, our subanalyses may have been biased toward the null effect, especially for SNPs with moderate effect sizes, due to a lack of power. Our study also lacked an independent cohort for validation of our findings due to the rarity of the disease.

We demonstrated that common germline variants are associated with EFS and OS among individuals with RMS. This GWAS provides evidence to support future investigation into the biological mechanisms underlying these associations. With replication of our findings and further understanding of the role of these SNPs in RMS-specific contexts, common germline variants might be of prognostic value for RMS diagnostic risk stratification.

Data availability

The data underlying this article are available through dbGaP (accession number phs003192.v1).

Author contributions

Funding

Conflicts of interest

The authors have no conflicts of interest to disclose.

Acknowledgements

References

1. Hibbitts E, Chi YY, Hawkins DS, et al. Refinement of risk stratification for childhood rhabdomyosarcoma using FOXO1 fusion status in addition to established clinical outcome predictors: a report from the Children's Oncology Group. *Cancer Med*. 2019;8(14):6437-6448.

2. Skapek SX, Anderson J, Barr FG, et al. PAX-FOXO1 fusion status drives unfavorable outcome for children with rhabdomyosarcoma: a Children's Oncology Group report. *Pediatr Blood Cancer*. 2013;60(9):1411-1417.

3. Williamson D, Missiaglia E, de Reynies A, et al. Fusion gene-negative alveolar rhabdomyosarcoma is clinically and molecularly indistinguishable from embryonal rhabdomyosarcoma. *J Clin Oncol*. 2010;28(13):2151-2158.

4. Haduong JH, Heske CM, Allen-Rhoades W, et al. An update on rhabdomyosarcoma risk stratification and the rationale for current and future Children's Oncology Group clinical trials. *Pediatr Blood Cancer*. 2022;69(4):e29511.

5. Malempati S, Weigel BJ, Chi YY, et al. The addition of cixutumumab or temozolomide to intensive multiagent chemotherapy is feasible but does not improve outcome for patients with metastatic rhabdomyosarcoma: a report from the Children's Oncology Group. *Cancer*. 2019;125(2):290-297.

6. 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*. 2018;36(27):2770-2777.

7. Weigel BJ, Lyden E, Anderson JR, et al. Intensive multiagent therapy, including dose-compressed cycles of ifosfamide/etoposide and vincristine/doxorubicin/cyclophosphamide, irinotecan, and radiation, in patients with high-risk rhabdomyosarcoma: a report from the Children's Oncology Group. *J Clin Oncol*. 2016;34(2):117-122.

8. Shern JF, Selfe J, Izquierdo E, et al. Genomic classification and clinical outcome in rhabdomyosarcoma: a report from an international consortium. *J Clin Oncol*. 2021;39(26):2859-2871.

9. Martin-Giacalone BA, Weinstein PA, Plon SE, Lupo PJ. Pediatric rhabdomyosarcoma: epidemiology and genetic susceptibility. *J Clin Med*. 2021;10(9):2028.

10. Koster R, Panagiotou OA, Wheeler WA, et al. Genome-wide association study identifies the GLDC/IL33 locus associated with survival of osteosarcoma patients. *Int J Cancer*. 2018;142(8):1594-1601.

11. Lin Q, Han J, Sun Q, et al. Functional variant of IL33 is associated with survival of osteosarcoma patients. *J Bone Oncol*. 2020;20:100270.

12. Perez-Andreu V, Roberts KG, Harvey RC, et al. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nat Genet*. 2013;45(12):1494-1498.

13. Zhang H, Liu AP, Devidas M, et al. Association of GATA3 polymorphisms with minimal residual disease and relapse risk in childhood acute lymphoblastic leukemia. *J Natl Cancer Inst*. 2021;113(4):408-417.

14. Rudzinski ER, Anderson JR, Hawkins DS, et al. The World Health Organization classification of skeletal muscle tumors in pediatric rhabdomyosarcoma: a report from the Children's Oncology Group. *Arch Pathol Lab Med*. 2015;139(10):1281-1287.

15. Das S, Forer L, Schonherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287.

16. Rizvi AA, Karaesmen E, Morgan M, et al. gwasurvivr: an R package for genome-wide survival analysis. *Bioinformatics*. 2019;35(11):1968-1970.

17. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.

18. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nat Genet*. 2004;36(5):512-517.

19. International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;437(7063):1299-1320.

20. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76-82.

21. Boughton AP, Welch RP, Flickinger M, et al. LocusZoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics*. 2021;37(18):3017-3018. doi:10.1093/bioinformatics/btab186.

22. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585.

23. Ognjanovic S, Linabery AM, Charbonneau B, et al. Trends in childhood rhabdomyosarcoma incidence and survival in the United States, 1975-2005. *Cancer*. 2009;115(18):4218-4226.

24. Sultan I, Qaddoumi I, Yaser S, et al. Comparing adult and pediatric rhabdomyosarcoma in the surveillance, epidemiology and end results program, 1973 to 2005: an analysis of 2,600 patients. *J Clin Oncol*. 2009;27(20):3391-3397.

25. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. 2012;40(Database issue):D930-D934.

26. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797.

27. Khan J. OncoGenomics DB [dataset]. National Cancer Institute. 2022. https://omics-oncogenomics.ccr.cancer.gov/cgi-bin/JK. Accessed February 1, 2023.

28. Hagenbuch B, Gui C. Xenobiotic transporters of the human organic anion transporting polypeptides (OATP) family. *Xenobiotica*. 2008;38(7-8):778-801.

29. Lopez-Lopez E, Martin-Guerrero I, Ballesteros J, et al. Polymorphisms of the SLCO1B1 gene predict methotrexate-related toxicity in childhood acute lymphoblastic leukemia. *Pediatr Blood Cancer*. 2011;57(4):612-619.

30. Ramsey LB, Bruun GH, Yang W, et al. Rare versus common variants in pharmacogenetics: SLCO1B1 variation and methotrexate disposition. *Genome Res*. 2012;22(1):1-8.

31. Trevino LR, Shimasaki N, Yang W, et al. Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J Clin Oncol*. 2009;27(35):5972-5978.

32. Sakaguchi S, Garcia-Bournissen F, Kim R, et al. Prolonged neutropenia after irinotecan-based chemotherapy in a child with polymorphisms of UGT1A1 and SLCO1B1. *Arch Dis Child*. 2009;94(12):981-982.

33. Nozawa T, Minami H, Sugiura S, et al. Role of organic anion transporter OATP1B1 (OATP-C) in hepatic uptake of irinotecan and its active metabolite, 7-ethyl-10-hydroxycamptothecin: in vitro evidence and effect of single nucleotide polymorphisms. *Drug Metab Dispos*. 2005;33(3):434-439.

34. Bao Z, Zhu R, Fan H, Ye Y, Li T, Chai D. Aberrant expression of SPAG6 and NM23 predicts poor prognosis of human osteosarcoma. *Front Genet*. 2022;13:1012548.

35. Neesse A, Gangeswaran R, Luettges J, et al. Sperm-associated antigen 1 is expressed early in pancreatic tumorigenesis and promotes motility of cancer cells. *Oncogene*. 2007;26(11):1533-1545.

36. Giannos P, Kechagias KS, Gal A. Identification of prognostic gene biomarkers in non-small cell lung cancer progression by integrated bioinformatics analysis. *Biology (Basel)*. 2021;10(11):1200.

37. Vite A, Li J, Radice GL. New functions for alpha-catenins in health and disease: from cancer to heart regeneration. *Cell Tissue Res*. 2015;360(3):773-783.

38. Shukla N, Ameur N, Yilmaz I, et al. Oncogene mutation profiling of pediatric solid tumors reveals significant subsets of embryonal rhabdomyosarcoma and neuroblastoma with mutated genes in growth signaling pathways. *Clin Cancer Res*. 2012;18(3):748-757.

39. Kobielak A, Fuchs E. Alpha-catenin: at the junction of intercellular adhesion and actin dynamics. *Nat Rev Mol Cell Biol*. 2004;5(8):614-625.

40. Jeronimo C, Robert F. The mediator complex: at the nexus of RNA polymerase II transcription. *Trends Cell Biol*. 2017;27(10):765-783.

41. Jiang C, Chen H, Shao L, et al. MicroRNA-1 functions as a potential tumor suppressor in osteosarcoma by targeting Med1 and Med31. *Oncol Rep*. 2014;32(3):1249-1256.

42. Schiano C, Rienzo M, Casamassimi A, et al. Gene expression profile of the whole Mediator complex in human osteosarcoma and normal osteoblasts. *Med Oncol*. 2013;30(4):739.

43. Schiano C, Casamassimi A, Rienzo M, de Nigris F, Sommese L, Napoli C. Involvement of mediator complex in malignancy. *Biochim Biophys Acta*. 2014;1845(1):66-83.

44. Malempati S, Hawkins DS. Rhabdomyosarcoma: review of the Children's Oncology Group (COG) Soft-Tissue Sarcoma Committee experience and rationale for current COG studies. *Pediatr Blood Cancer*. 2012;59(1):5-10.