SUPPLEMENTARY MATERIALS

Germline cancer-predisposition variants in pediatric rhabdomyosarcoma: a report from the Children's Oncology Group

He Li, PhD[1], Saumya D. Sisoudiya, BS[2,3,4], Bailey A. Martin-Giacalone, BS[2], Michael M. Khayat, BS[1,3], Shannon Dugan-Perez, BS[1], Deborah A. Marquez-Do, BS[2], Michael E. Scheurer, PhD[2,4], Donna Muzny, MS[1], Eric Boerwinkle, PhD[1,5], Richard A. Gibbs, PhD[1,3], Yueh-Yun Chi, PhD[6], Donald A. Barkauskas, PhD[7,8], Tammy Lo, MPH[7], David Hall, MS[7], Douglas R. Stewart, MD[9], Joshua D. Schiffman, MD[10], Stephen X. Skapek, MD[11], Douglas S. Hawkins, MD[12], Sharon E. Plon, MD[1,2,3], Aniko Sabo, PhD[1*], Philip J. Lupo, PhD[2,4*]

1. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2. Section of Hematology-Oncology, Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 4. Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX; 5. School of Public Health, the University of Texas Health Science Center, Houston, TX; 6. Children's Hospital Los Angeles, University of Southern California, Los Angeles, California; 7. QuadW Childhood Sarcoma Biostatistics and Annotation Office at the Children's Oncology Group, Monrovia, CA; 8. Department of Preventive Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA; 9. Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 10. Departments of Pediatrics and Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 11. Department of Pediatrics, the University of Texas Southwestern Medical Center, Dallas, TX; 12. Division of Hematology-Oncology, Seattle Children's Hospital, University of Washington, Seattle, WA. *Equally contributing senior authors

Correspondence to:

Philip J. Lupo, Ph.D.

Department of Pediatrics, Baylor College of Medicine

6620 Main Street, BCLN-11D.26, Houston, TX, 77030

E-mail: philip.lupo@bcm.edu

Tel: (713) 798-2960

Supplementary Methods

Exome sequencing and variant filtering

Exome sequencing of the RMS samples was performed at the Human Genome Sequencing Center (HGSC) at Baylor College of Medicine. Specifically, genomic DNA samples underwent exome capture using HGSC VCRome2.1 (35Mb covering ~24K genes; Roche) and were then sequenced on the Illumina NovaSeq platform with ~100X coverage. Alignment was performed using BWA-MEM (version 0.7.15-r1140) (1) against the human hs37d5 assembly, followed by Base Score Recalibration and Indel Realignment using GATK (version 3.4-0) (2). Variant calling was performed using xAtlas (version 0.2.1) (3) to generate gvcf files for merging different sets of data. We did not call larger deletions in this study. Samples from the ARIC and VIVA cohorts have been sequenced at the HGSC on the same platform and described previously (4,5). Of note, both studies have used similar exome capture reagent (VCrome2.1) and were subjected to the same analytical pipelines as the RMS cohort.

Next, we performed joint genotyping using GLnexus (v1.1.3) (6) to combine samples from the three datasets (RMS patients, ARIC controls, and VIVA controls) to increase consistency of variant calling. A series of quality control (QC) measures were then employed. First, for each variant in each sample, we assigned a missing genotype if the variant fulfilled one of the following criteria: read depth less than 15; quality score less than 15; allelic ratio less than 0.25 or greater than 0.75; or non-PASS variants from xAtlas variant call results. We then removed the variants with a genotyping rate less than 85% within each dataset. These QC measures were performed independently for single nucleotide variants (SNVs) and insertion/deletions (INDELs). For all the pathogenic variants found in RMS patients, we manually reviewed each of them in IGV (V2.3) (7) to validate the variant. In addition, we performed sample-wise QC by removing those samples with a missing genotyping rate >10%. We also removed samples with a heterozygosity rate >3 standard deviations to

eliminate potential sample contamination. For each set of the cohorts, we calculated individual-wise identity by descent (IBD) using PLINK (v1.90b2t) (8) and kept unrelated samples with IBD <0.1 via PRIMUS (v1.9.0) (9). None of the RMS samples were related based on the above criteria. Genetic ancestry of each individual was determined using ADMIXTURE (v1.3.0) (10). After filtering, we included 615 RMS patients (99.7% of variants > 30X depth), 9,663 controls from ARIC (97.8% of variants > 30X depth), and 300 controls from VIVA (99.0% of variants > 30X depth) in the final analyses.

Curation of cancer-predisposition genes (CPGs)

We pre-curated a set of autosomal dominant CPGs (n = 24) from cancer-susceptibility syndromes that have specifically been implicated in RMS predisposition including Li-Fraumeni (11), neurofibromatosis type 1 (12,13), Costello (14,15), DICER1 (16), Gorlin (17), Noonan (14), Cardiofaciocutaneous (14), constitutional (biallelic) mismatch repair deficiency (18), Beckwith-Wiedemann (19), and Rubinstein-Taybi (20). Notably, 21 of these 24 RMS-associated CPGs were included in the assessment by Zhang *et al.* (21) for the prevalence of germline pathogenic variants in other cancers. We also included a set of additional autosomal dominant CPGs (n = 39) from the report by Zhang *et al.* (21).

We also conducted a separate analysis to evaluate the frequency of pathogenic variants in autosomal recessive genes associated with sarcomas. We searched the "Online Mendelian Inheritance in Man" catalog using the search terms "rhabdomyosarcoma" and "sarcoma" to identify genes implicated in autosomal recessive disorders associated with sarcomas and selected the seven following genes: *RECQL2*, *RECQL4*, *BUB1B*, *NBN*, *CEP57*, *TRIP13*, and *BLM*.

Supplementary References

1. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013;1303.3997v2.

2. McKenna A, Hanna M, Banks E*, et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303.

3. Farek J, Hughes D, Mansfield A*, et al.* xAtlas: Scalable small variant calling across heterogeneous next-generation sequencing experiments. *bioRxiv*. 2018;doi: https://doi.org/10.1101/295071.

4. Butte NF, Cai G, Cole SA, Comuzzie AG. Viva la Familia Study: genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. *Am J Clin Nutr*. 2006;84(3):646-654; quiz 673-644.

5. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol*. 1989;129(4):687-702.

6. Lin MF, Rodeh O, Penn J*, et al.* GLnexus: joint variant calling for large cohort sequencing. *bioRxiv*. 2018;doi: https://doi.org/10.1101/343970.

7. Robinson JT, Thorvaldsdottir H, Winckler W*, et al.* Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26.

8. Purcell S, Neale B, Todd-Brown K*, et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575.

9. Staples J, Nickerson DA, Below JE. Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. *Genet Epidemiol*. 2013;37(2):136-141.

10. Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*. 2011;12:246.

11. Diller L, Sexsmith E, Gottlieb A, Li FP, Malkin D. Germline p53 mutations are frequently detected in young children with rhabdomyosarcoma. *J Clin Invest*. 1995;95(4):1606-1611.

12. Hartley AL, Birch JM, Marsden HB, Harris M, Blair V. Neurofibromatosis in children with soft tissue sarcoma. *Pediatr Hematol Oncol*. 1988;5(1):7-16.

13. Yang P, Grufferman S, Khoury MJ*, et al.* Association of childhood rhabdomyosarcoma with neurofibromatosis type I and birth defects. *Genet Epidemiol*. 1995;12(5):467-474.

14. Kratz CP, Rapisuwon S, Reed H, Hasle H, Rosenberg PS. Cancer in Noonan, Costello, cardiofaciocutaneous and LEOPARD syndromes. *Am J Med Genet C Semin Med Genet*. 2011;157C(2):83-89.

15. Estep AL, Tidyman WE, Teitell MA, Cotter PD, Rauen KA. HRAS mutations in Costello syndrome: detection of constitutional activating mutations in codon 12 and 13 and loss of wild-type allele in malignancy. *Am J Med Genet A*. 2006;140(1):8-16.

16. Doros L, Yang J, Dehner L*, et al.* DICER1 mutations in embryonal rhabdomyosarcomas from children with and without familial PPB-tumor predisposition syndrome. *Pediatr Blood Cancer*. 2012;59(3):558-560.

17. Hahn H, Wojnowski L, Zimmer AM*, et al.* Rhabdomyosarcomas and radiation hypersensitivity in a mouse model of Gorlin syndrome. *Nat Med*. 1998;4(5):619-622.

18. Kratz CP, Holter S, Etzler J*, et al.* Rhabdomyosarcoma in patients with constitutional mismatch-repair-deficiency syndrome. *J Med Genet*. 2009;46(6):418-420.

19. Mussa A, Molinatto C, Baldassarre G*, et al.* Cancer Risk in Beckwith-Wiedemann Syndrome: A Systematic Review and Meta-Analysis Outlining a Novel (Epi)Genotype Specific Histotype Targeted Screening Protocol. *J Pediatr*. 2016;176:142-149 e141.

20. Stevens CA. Rubinstein-Taybi Syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al., (eds). *GeneReviews((R))*. Seattle (WA); 1993.

21. Zhang J, Walsh MF, Wu G*, et al.* Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med*. 2015;373(24):2336-2346.

Supplementary Tables

Supplementary Table 1. Epidemiological features of cases and controls used in this study

		Controls			RMS cases
		ARIC	VIVA	Total	
Number of samples		9,663	300	9,963	615
Sex	Female	5,417 (56.1%)	163 (54.3%)	5,580 (56.0%)	243 (39.5%)
	Male	4,246 (43.9%)	137 (45.7%)	4,383 (44.0%)	372 (60.5%)
Population	European American	7,122 (73.7%)	0	7,122 (71.5%)	405 (65.9%)
	African American	2,541 (26.3%)	0	2,541 (25.5%)	78 (12.7%)
	Hispanic	0	300 (100%)	300 (3.0%)	96 (15.6%)
	East Asian	0	0	0	28 (4.6%)
	South Asian	0	0	0	8 (1.3%)
Age* (year)	0 - 10	0	127 (42.3%)	127 (1.3%)	404 (65.7%)
	10- 20	0	146 (48.7%)	146 (1.5%)	198 (32.2%)
	20 - 30	0	0	0	13 (2.1%)
	30 - 40	0	0	0	0
	40 - 50	1,979 (20.5%)	0	1,979 (19.9%)	0
	50 - 60	2,986 (30.9%)	0	2,986 (30.0%)	0
	> 60	1,114 (11.5%)	0	1,114 (11.1%)	0
	NA	3,584 (37.1%)	27 (9.0%)	3,611 (36.2%)	0

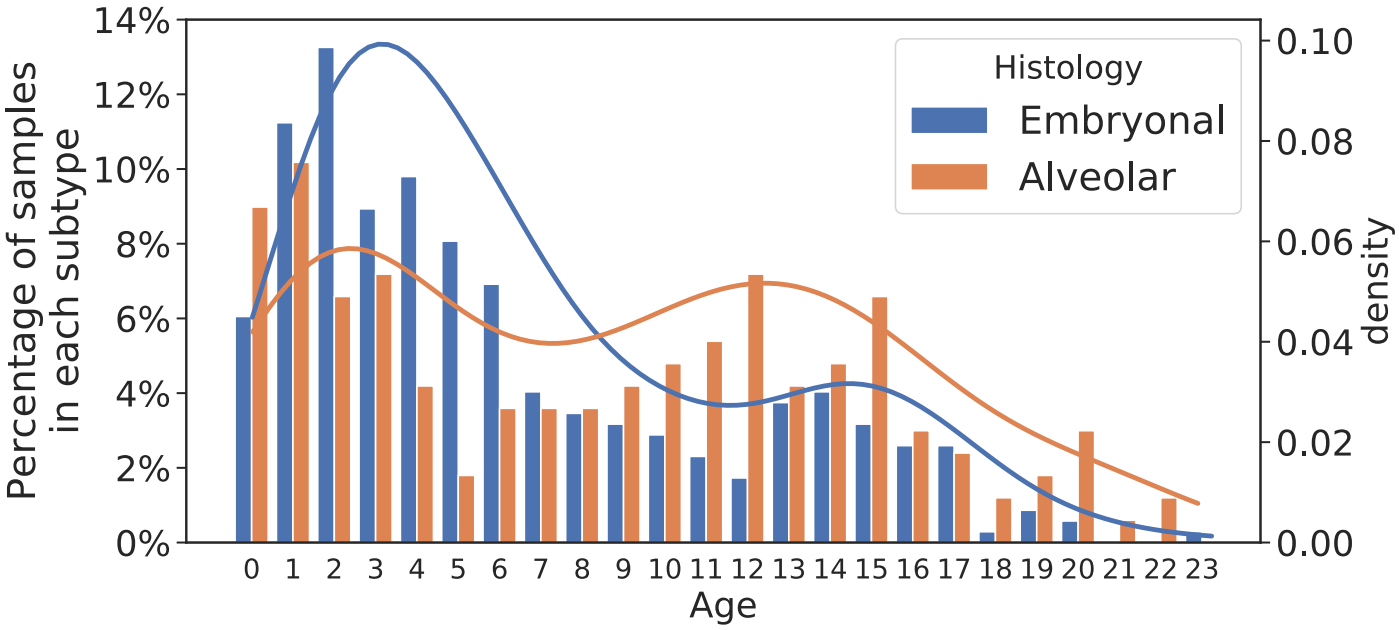
* For controls, Age represents the time when the sequencing experiment was performed. For RMS cases, Age represents the time when a diagnosis was made.

Supplementary Table 2. Associations between pathogenic or likely pathogenic (P/LP) variant frequencies and case / control status in different populations

	RMS cases		Controls		p value
	Total	with P/LP	Total	with P/LP	OR (95% CI)*
European American	405	33 (8.2%)	7122	96 (1.4%)	2.20E-19 7.27 (4.72-11.20)
African American	78	5 (6.4%)	2541	47 (1.9%)	5.96E-3 3.97 (1.49-10.60)
Hispanic	96	5 (5.2%)	300	3 (1.0%)	3.27E-3 13.2 (2.37-73.72)
Chi-square tests across populations	$p = 0.57$; $\chi^2 = 1.11$		$p = 0.16$; $\chi^2 = 3.73$		

*The control groups were served as the reference group for calculating odds ratio

Supplementary Figures



Supplementary Figure 1. Age distributions for embryonal and alveolar rhabdomyosarcoma patients. Bar plots indicate the percentage of patients of a given age from each histology group.