Multiple new Ewing sarcoma susceptibility loci expand knowledge of germline genetic etiology and nominate mechanisms of risk

Authors: Aubrey K. Hubbard[1], Weiyin Zhou[1], Carrie Boyce[2], Javed Khan[3], Melissa M. Hudson[4], Kirsten K. Ness[4], Zhaoming Wang[5], Katherine A. Janeway[6], Philip J. Lupo[7], Lindsay M. Morton[1], Stephen J. Chanock[1], Thomas G.P. Grünewald[8,9,10,11], Olivier Delattre[12], Mitchell J. Machiela[1] on behalf of the Ewing Sarcoma Germline Genetics Group

1. Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA
2. Westat, Rockville, MD, USA
3. Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA
4. Department of Oncology, St. Jude Children's Research Hospital, Memphis,TN, USA
5. Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA
6. Dana-Farber/Boston Children's Cancer and Blood Disorders Center, Boston, MA, USA.
7. Department of Pediatrics, Section of Hematology-Oncology, Baylor College of Medicine, Houston, TX, USA
8. Division of Translational Pediatric Sarcoma Research, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), 69120 Heidelberg, Germany.
9. Hopp-Children's Cancer Center (KiTZ), 69120 Heidelberg, Germany.
10. Institute of Pathology, Heidelberg University Hospital, 69120 Heidelberg, Germany
11. National Center for Tumor Diseases (NCT), NCT Heidelberg, a partnership between DKFZ and Heidelberg University Hospital, Germany
12. INSERM U830, Diversity and Plasticity of Childhood Tumors Laboratory, PSL Research University, SIREDO Oncology Center, Institut Curie Research Center, Paris 75005, France

---

Ewing sarcoma (EwS) is a rare and aggressive pediatric cancer with peak incidence in the second decade of life. EwS occurs primarily in children of European ancestry suggesting a genetic component to susceptibility. Prior investigations identified six susceptibility loci associated with EwS, with evidence for additional signals not reaching the commonly accepted threshold for genome-wide significance (i.e., $p<5\times10^{-8}$). At least two of these loci demonstrate evidence of tagging variation in length of nearby GGAA microsatellites, which alter binding of the pathognomonic *EWSR1::FLI1* fusion resulting in downstream dysregulation of target genes promoting sarcomagenesis. To identify additional susceptibility loci, we performed the largest genome-wide association study (GWAS) to date consisting of 1,640 EwS cases and 8,457 cancer-free individuals of European ancestry from 8 studies and matched by principal components. Genotype imputation was carried out by study/array using TOPMed data as the reference and each set was analyzed using PLINK 1.9 adjusting for significant principal

components. Variants were included in each study if minor allele count was greater than five in each set of cases. Individual study results were combined using fixed effects meta-analysis in Metasoft with variants demonstrating evidence for heterogeneity excluded.

For previously reported EwS loci (i.e., 1p36.22, 6p25.1, 10q21.3, 15q15.1, 20p11.22 and 20p11.23), odds ratios (OR) remained high for GWAS (OR=1.5-2.4) and the $-\log_{10}$ p-values grew by several orders of magnitude, providing strong evidence for replication in the additional 907 EwS cases and 7,111 cancer-free controls. Our analysis identified five novel loci reaching genome-wide significance ($P\text{-value}_{meta} < 5\times10^{-8}$) at 5q32, 7q32.3, 8q24.2, 13q32.3, and 17q23.2. Minor allele frequency of newly identified loci ranged from 0.001 to 0.36 with estimated ORs ranging from 1.31-4.30. eQTL analyses nominated nearby biologically plausible candidate genes for future functional investigation including *LARS1*, a gene implicated in osteosarcoma and skeletal muscle dysgenesis, at the 5q32 locus and *COX5BP6,* a gene implicated in age at menarche and testosterone levels, at 13q32.3.

Our results add to evidence supporting a strong inherited genetic component to EwS risk and particularly a genetic architecture harboring a substantial number of common variants with moderate effect. Ongoing work is focusing on enrichment analyses of GGAA microsatellite repeat sequences in identified risk loci to search for evidence for germline-somatic interactions with *EWSR1::FLI1*.