Guidelines: You are allotted 2,300 characters, excluding spaces, for the body of your abstract. Title, author, and institutional data are not included in the character limit.

Targeted long-read sequencing of the Ewing sarcoma 6p25.1 susceptibility locus identifies polymorphic GGAA microsatellite associated with EWSR1-FLI1 fusion binding

Olivia Lee[1,2], Shu-Hong Lin[1], Wen Luo[1,3], Kristine Jones[1,3], Derek W. Brown[1], Robert Hoover[1], Gregory T. Armstrong, Smita Bhatia, Leslie L. Robison, Neal Freedman[1], Lindsay Morton[1], Stephen Chanock[1], Olivier Delattre[7,8], Thomas Grunewald[9,10,11], and Mitchell Machiela[1]

(1) Division on Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD.
(2) Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX.
(3) Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Leidos Biomedical Research Inc, Frederick, MD.
(4) Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN
(5) Institute for Cancer Outcomes and Survivorship, University of Alabama, Birmingham, AL
(6) Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN
(7) Inserm U830, Équipe Labellisés LNCC, PSL Université, Institut Curie, Paris, France
(8) SIREDO Oncology Centre, Institut Curie, Paris, France
(9) Max-Eder Research Group for Pediatric Sarcoma Biology, Ludwig Maximilians Universität (LMU), Munich, Germany
(10) German Cancer Consortium (DKTK), partner site Munich, Munich, Germany
(11) German Cancer Research Center (DKFZ), Heidelberg, Germany

Topics: 12 – Statistical Genetics and Genetic Epidemiology
Subtopic: Cancer
Five Keywords: Cancer, Genetic epidemiology, Genome-wide association, Linkage disequilibrium, Susceptibility locus

Ewing sarcoma (EwS) is the second most commonly diagnosed bone malignancy in children and adolescents. EwS is driven by a balanced chromosomal translocation between a member of the FET protein family (usually *EWSR1*) and a gene in the ETS family (usually *FLI1*), resulting in the production of a novel chimeric oncoprotein that binds to GGAA DNA motifs and substantially rewires nearby transcriptional regulation. GWAS of EwS have also identified 6

germline susceptibility loci associated with EwS risk, all with high associated odds ratios (ORs>1.7). Despite the rarity of EwS, tagging variants in EwS susceptibility loci are common (MAF>5%) and often reside near GGAA microsatellite sequences, suggesting germline-somatic interactions are key elements in the genetic architecture of EwS. We performed targeted, long read PacBio sequencing on 271 EwS cases and 114 cancer-free adult controls at chromosome 6p25.1 to comprehensively characterize germline variation in this EwS susceptibility region. We profiled alleles of a nearby microsatellite and phased with local SNPs to disentangle the relationship of germline variation and EwS risk at this locus.

We detected 50 unique microsatellites alleles and observed EwS cases on average had a higher proportion of long microsatellites alleles compared to controls (17.32% versus 11.43%; p=0.016). EwS cases had microsatellite alleles with a higher number of GGAA motif repeats compared to controls (19.25 versus 18.51; p=0.009) suggesting that GGAA motif content is a major contributor linking elongated microsatellite length at 6p25.1 to increased EwS risk. Other non-GGAA motifs present in the targeted 6p25.1 microsatellite sequence (e.g., AGAA and GGGA motifs) were not associated with EwS risk. In addition, we observed longer microsatellites were highly correlated with the risk allele (A) of rs7742053, the lead 6p25.1 variant in our EwS GWAS. These findings indicate that the rs7742053 EwS GWAS variant tags longer GGAA microsatellite repeat alleles.

ChIP-seq data indicated the sequenced microsatellite at 6p25.1 is in a region of open chromatin and that FLI1 binding also occurs. Furthermore, expression quantitative trait locus analysis identified *RREB1* (p=0.01), a RAS responsive element, as a candidate gene at 6p25.1 likely dysregulated by EWSR1-FLI1 binding. Expression analysis across sarcomas reveals EwS tumors have elevated *RREB1* expression. Our comprehensive long-read sequencing of germline variation in the 6p25.1 EwS locus suggests increased GGAA repeats in the polymorphic microsatellite is associated with elevated EwS risk and may interact with somatic EWSR1-FLI1 fusions to dysregulate *RREB1* expression.

Characters (without spaces): 2,288