**Title**: Rare high-penetrance and common low-penetrance variants associated with risk of pediatric acute lymphoblastic leukemia

**Authors**: Cheng Chen, Na Qin, Nan Song, Hui Wang, Gregory T. Armstrong, Kirsten K. Ness, Melissa M. Hudson, Leslie L. Robison, Cindy Im, Zhaoming Wang

**Abstract**: Although there has been a remarkable progress in cancer genetics research, knowledge of inherited genetic risk factors for pediatric acute lymphoblastic leukemia (ALL) is incomplete. Leveraging existing genetic data for survivors from the St. Jude Lifetime Cohort Study (SJLIFE) and Childhood Cancer Survivor Study (CCSS), genetic variants across full allelic spectrum were analyzed for their associations with ALL risk. A total of 2,695 ALL cases with whole-genome or whole-exome sequencing were available for rare variant analyses. Pathogenic/likely pathogenic variants in 60 genes associated with autosomal dominant cancer predisposition syndromes were characterized and classified with "PeCanPIE" - the Pediatric Cancer Variant Pathogenicity Information Exchange, a web- and cloud-based platform. We subsequently carried out a rare variant enrichment analysis by comparing ALL survivors with a general population sample from the GnomAD database (n~60,000) and found seven statistically significant genes after multiple testing adjustment ($P< 8.3x10^{-4}$, i.e., 0.05/60) including: *BRCA1* (odds ratio [OR], 4.06; 95% CI, 2.00-7.59; P, $1.21\times10^{-4}$), *BRCA2* (OR, 3.48; 95% CI, 1.82-6.15; P, $1.50\times10^{-4}$), *CDKN2A* (OR, 22.30; 95% CI, 5.13-96.68; P, $3.07\times10^{-5}$), *PALB2* (OR, 5.12; 95% CI, 2.39-10.00, P, $3.77\times10^{-5}$), *PAX5* (OR, 44.58; 95% CI, 6.49-490.68; P, $4.77\times10^{-5}$), *PTPN1* (OR, 66.85; 95% CI, 11.96-670.57, P, $1.63\times10^{-7}$), and *TP53* (OR, 17.86; 95% CI, 6.12-50.36, P, $3.43\times10^{-7}$). Repeating these analyses with 309 newly diagnosed patients from the Pediatric Cancer Genome Project, we observed similar results for five of these genes. To study common variants associated with ALL risk, we performed a genome-wide association analysis using 2,777 ALL survivor cases and 6,255 controls (other pediatric cancer cases and non-cancer community controls) with whole-genome sequencing or imputed SNP-array data. In addition to replicating known associations at *ARID5B*, *IKZF1*, *CDKN2A*, *BMI1*, *PIP4K2A*, *CCDC26* and *CEPBE*, we identified a novel locus (rs112425636, chr17:82324152:G:A; OR, 1.65; 95% CI, 1.42-1.93, P, $2.48\times10^{-10}$) mapped to the intronic region of *SECTM1* gene. Based on RNA sequencing data for 164 pediatric hematological cancer samples, we found that the RNA expression of *SECTM1* gene was significantly higher among patients with GG genotype than patients with AA or AG genotypes for rs112425636 (FPKM=1.34 vs. 0.71, P<0.01). Among the subset of 32 B-ALL samples, the magnitude of difference was larger (FPKM=1.90 vs. 0.70, P=0.017). Research using cancer survivors to investigate cancer etiology is subject to potential survival bias, but pediatric ALL survival rates are high. Moreover, our replication of previous findings affirms this a reasonable approach. In summary, we found statistical evidence for several cancer predisposition genes harboring rare high-penetrance variants and a novel common low-penetrance variant associated with risk of pediatric ALL. Our novel findings add new knowledge to the full allelic spectrum of genetic architecture of pediatric ALL susceptibility.