Recurrent Repeat Contractions: Investigating a Novel Genomic Factor of Polymorphism in 10 Human Cancers

Zhu, Kevin (School: Jericho High School)

Replication errors represent two-thirds of cancer mutations. A novel replication error subset, recurrent repeat contractions (rRCs), lack systematic examination in cancers due to complex amplifications and intratumor heterogeneity. This study aimed to determine if rRCs are significant cancer mutation sources by 1) identifying rRCs' prevalence in cancers, 2) creating a local read depth normalization algorithm (LRDN) to reduce false-positive rRCs, 3) comparing rRCs in cancer to pathogenic rRCs in other diseases and microsatellites in cancer, and 4) conducting a case-study on clinically-relevant rRCs. ExpansionHunter Denovo analysis on 2,622 genomes spanning 37 cancer types with filtering ($p<0.05$) identified 689 candidate rRCs. The developed LRDN algorithm removed >80% of false candidate rRCs, yielding rRCs (N=120) across 10 cancer types. Genomic qualities of cancer rRCs resembled rRCs in other diseases, including a bimodal motif length distribution, tendency towards telomeric and genic regions, and closer proximity to cis-regulatory elements than expected ($p=6.00e-45$), suggesting rRCs role in cell death and gene regulation. Shockingly, rRCs weren't correlated with microsatellite instability ($p=0.27$), demonstrating that rRCs may arise by a distinct pathway. Disease-gene associations of rRCs to human diseases revealed correlations with carcinoma ($p=3.45e-4$). A case study of Hepatocellular Carcinoma (HCC) revealed 46% contained a PRDM16-located intronic rRC associated with decreased isoform usage ($p=0.0048$), while high PRDM16 expression correlated to low HCC survival rates ($p=2.2x10-7$), suggesting a functional rRC in HCC. Future investigations require determination of causal mechanisms of rRCs in cancers, as contraction-based targets may allow for future cancer therapeutics.

Awards Won:

First Award of $5,000

First Award of $5,000