

Insights: Publications

Functional classification of long non-coding RNAs by k-mer content

Nature Genetics Volume 50, 1474–1482

September 17, 2018

The functions of most long non-coding RNAs (lncRNAs) are unknown. In contrast to proteins, lncRNAs with similar functions often lack linear sequence homology; thus, the identification of function in one lncRNA rarely informs the identification of function in others. We developed a sequence comparison method to deconstruct linear sequence relationships in lncRNAs and evaluate similarity based on the abundance of short motifs called k-mers. We found that lncRNAs of related function often had similar k-mer profiles despite lacking linear homology, and that k-mer profiles correlated with protein binding to lncRNAs and with their subcellular localization. Using a novel assay to quantify Xist-like regulatory potential, we directly demonstrated that evolutionarily unrelated lncRNAs can encode similar function through different spatial arrangements of related sequence motifs. K-mer-based classification is a powerful approach to detect recurrent relationships between sequence and function in lncRNAs.