Cis-regulatory G-quadruplex Motifs are Preferentially Associated with Splice Sites in the Protein-Coding Human Genome

! " # $! " %

April 2020

Abstract

Expression of mammalian genes involves regulated RNA splicing. Most human genes undergo alternative splicing during gene expression. As a result, the human protein-coding genome provides a rich variety of proteins ith complex and diverse functions. !t is estimated that up to one-fifth of human diseases are associated ith altered splicing.

"ur la# studies the role of cis-regulatory motifs, such as $uadruplex forming %-Rich &e'uences ($%R&) in RNA processing. *e focus on computationally identifying $%R& distri#ution patterns near splice sites in the human protein-coding genome and investigate their role in regulated splicing. "ur dataset consists of +,,,-. genes, /,0,--0 exons, //2,--0 introns, and -1,,,2 uni'ue splice sites #ased on the %R3h-. 4omo sapiens assem#ly extracted from the 4uman Ensem#l data#ase. *e have developed scripts in 5ython- and 366, #ased on our previously esta#lished $%R& Mapper program, to map $%R& motifs.

"ur analysis discovered a preferential association of $%R& motifs ith splice sites in exons