GC content provides new insights into exon recognition

Eddo Kim
kimedd@post.tau.ac.il

Amir Goren
amirgore@post.tau.ac.il

Maayan Amit
Maayana2@post.tau.ac.il

Gil Ast
gilast@post.tau.ac.il

1 Department of Human Molecular Genetics & Biochemistry, Sackler Faculty of Medicine, Tel-Aviv University, Israel.

# Contributed Equally.

Keywords: GC content, Exon definition, Intron definition

1 Introduction

Studies throughout the years revealed several characteristics that differentiate alternative exons from constitutive ones. However, some of those characteristics, such as splice signals and other regulatory sequences, are inherently sequence-based. Herein we show that these characteristics are heavily biased by GC content, and may not actually differ between alternative and constitutive exons. We further examine the effect of the GC content on exon versus introns definition. We show that the GC content environment affects the way exons and introns are recognized by the splicing machinery, namely different GC contents may dictate whether mRNAs are processed via exon or intron definition.

2 Methods and Results

Over the years, several studies revealed characteristics that differ between alternative and constitutive exons. Here we show that these studies were heavily biased by the GC content environment. One such characteristic found to differ in alternative and constitutive exons is the presence of ESEfinder (1) splicing regulatory motifs; they were found to be less prevalent in alternative versus constitutive exons (2). However, as splicing regulatory motifs are inherently sequence-based, we hypothesized that the large difference in GC content between different genes might strongly bias these findings. The results we present reveal that when accounting for GC content, there is virtually no difference in the presence of splicing regulatory sequences between alternative and constitutive exons (Figure 1).

The splice signals strength is another characteristic that was found to differ between constitutive and alternative exons (3). Because splicing signals also depend directly on the genomic sequences, we suspected that their strength, and especially the strength of the longer 3'ss which consists of the polypyrimidine tract, would depend greatly on the GC content environment. The results we show reveal that the difference in the splice signal strength is greater when comparing exons of different GC contents and similar mode of splicing, than comparing exons with a different mode of splicing but similar GC contents. Overall, accounting for GC content dramatically weakens the significance of the difference between the splice signal strength of constitutive and alternative exons (Figure 2-3).

Since GC content affects characteristics involved in exon recognition, we suspected that GC content may also directly affect exon recognition. Between 15% and 50% of mutations leading to genetic disorders activate cryptic splice sites or lead to exon skipping. Therefore, understanding the molecular mechanisms that differentiate between activation of cryptic splice site and exon skipping is of great importance. We hypothesized that the difference in GC content in different genes led to the appearance of two types of exon/intron structures, which the spliceosome handles differently. We further hypothesized that splicing-disrupting mutations adjacent to introns selected via the intron-definition mechanism activate aberrant splice sites, whereas similar mutations near exons selected via the exon-definition mechanism lead to exon skipping. We extracted 199 mutations that lead to exon skipping and 371 mutations that activate a cryptic splice site. Comparing the GC content of the environment surrounding these mutations revealed a significantly different landscape. We found a significantly higher level of GC content in near mutations that activate cryptic splice sites compared with mutations that resulted in exon skipping (Figure 4).
2.2 Figures

Figure 1: Analysis of ESEfinder motifs. The fraction of nucleotides that are part of an ESEfinder motifs was calculated. The fraction of exons (Y axis) that are covered by a specific fraction of ESEfinder motifs (X axis) is plotted for exons of high GC content (alternative: open triangles; constitutive: filled triangles), and for exons of low GC content (alternative: open squares; constitutive: filled squares).

Figure 2: Analysis of 5’ splice signal strength. The strength of the 5’ splice signal was calculated. The fraction of exons (Y axis) that exhibit a specific 5’ splice signal strength (X axis) is plotted for exons of high GC content (alternative: open triangles; constitutive: filled triangles), and for exons of low GC content (alternative: open squares; constitutive: filled squares).

Figure 3: Analysis of 3’ splice signal strength. The strength of the 3’ splice signal was calculated. The fraction of exons (Y axis) that exhibit a specific 3’ splice signal strength (X axis) is plotted for exons of high GC content (alternative: open triangles; constitutive: filled triangles), and for exons of low GC content (alternative: open squares; constitutive: filled squares).

Figure 4: Genomic areas harbor different types of mutations depending on the GC content environment. GC content was calculated (Y axis) for introns harboring mutations and their corresponding upstream and downstream exons and introns (X axis). The calculation was done separately for mutations that lead to activation of cryptic splices (striped squares) and for those that lead to exon skipping (dotted squares).

3 Discussion

Here we show that characteristics previously suggested as differentiating between alternative and constitutive exons are heavily biased by GC content, suggesting they may be only slightly linked, if at all, to alternative versus constitutive exon recognition. In most cases, even if these characteristics are indeed linked to the regulation of the mode of splicing, the magnitude of this link is substantially smaller than previously suggested, and the molecular mechanism that underlies this regulation is not straightforward. We further show that GC content affects the way exons and introns are recognized.

References