Characterization of minor intron and minor intron-containing genes (MIGs)

Beom Jin Park | Supervision: Ihab Younis
Carnegie. Mellon University Qatar, Biological Sciences 2020

Abstract

A minor intron, also known as U12 type intron, is distinguished by its unique dinucleotide 5' splice site (5'SS) and 3'SS. Conserved orthologously in the eukaryotic genome, minor intron has been spotlighted as an imperative biological component in our system. Here, we conducted statistical analysis on minor introns using online information (MIDB and NCBI RefSeq). Our analysis shows that minor introns are mostly distributed in chromosome 19, evenly distributed in forward and reverse strand, have average length of 4,627 (nt) and a median of 1292 nt, evenly distributed in a transcript, mostly possess pre-mature stop codon (PTC) which is most often positioned near the 3' end of minor intron. Our bioinformatics analysis will be conducive to understand the molecular mechanism of minor intron splicing regulation.

Background

Minor intron, also known as U12-type introns, is a rare group of introns that were initially identified based on their unique dinucleotides at the ends of the 5' and 3' splice sites: AT-AC (Turunen J. et al, 2013).While majority of introns (99.5% of all introns) are spliced out by major spliceosome composed of U1, U2, U4, U5 and U6 small nuclear ribonucleoproteins (snRNPs) , the minor introns are spliced by a similar yet different set of snRNPs, later depicted as the minor spliceosome (Wahl et al, 2009). The minor spliceosome is composed of U11, U12, U5, U4atac and U6atac, which are less abundant than their analogs: U1 (for U11), U2 (for U12), U4 (for U4atac) and U6 (for U6atac) (Turunen J. et al, 2013; Tarn et al, 1996). Previous studies show that minor intron splicing is inefficient compared to the major intron splicing and that such inefficiency in splicing leads to high retention rate of minor intron and disrupts normal level of gene expression Across eukaryotic species, minor introns show orthologous conservation, which hints their indispensable role in eukaryotic biological system. Recent finding shows that dysregulation in minor intron splicing leads to development of number of severe diseases such as MOPD1, Roifman syndrome, myelodysplastic syndrome and Lowry Wood syndrome (He et al., 2011; Merico et al., 2015, Farach et al., 2018; Maden et al., 2015)

Materials & Method

Bioinformatics analysis
Classification of minor introns
All information listed in MIGdb v.1.0 has been collected leveraging the information gathered from NCBI RefSeq (NCBI Annotation Release 109.20190905) and MIDB (Released 2019 from Kanadia Lab, Department of Physiology Neurobiology). The following sets of data were determined based on the longest isoform: minor intron number, exon count, minor intron position, downstream exon coordinate, upstream exon coordinate, and lost and gained sequence. If the protein-coding transcript was not present, then the longest non-protein coding transcript was used to determine the information.

Statistical analysis of minor intron
Minor intron distribution in chromosome
The minor intron distribution in human chromosome is normalized by the size of the chromosome. The size of the chromosome was identified from NCBI GenBank based on genome assembly, GRCH38.p13.
Minor intron distribution in DNA strand
The strand type of randomized major introns 696 minor introns were identified from NCBI RefSeq (NCBI Annotation Release 109.20190905) and compared.
Minor intron length distribution
The length of total U2 type introns in both protein-coding and non-coding genes in human genome were identified from NCBI RefSeq (NCBI Annotation Release 109.20190905). The length of minor introns was provided from MIDB.
Minor intron position distribution
The minor intron position was determined using normalizing minor intron number with total number of exons in a transcript. The minor intron number and total number of exons were determined from the longest isoform.

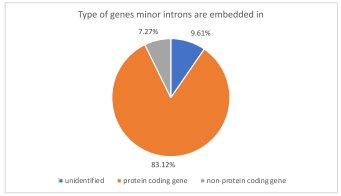

Figure 1. The gene types of MIGs

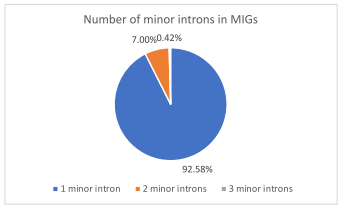

Figure 2. Number of minor introns in MIGs

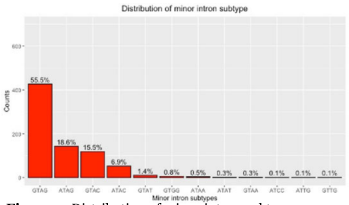

Figure 3. Distribution of minor intron subtype

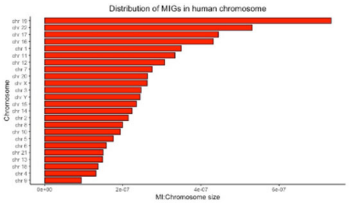

Figure 4. Distribution of minor introns in human chromosome

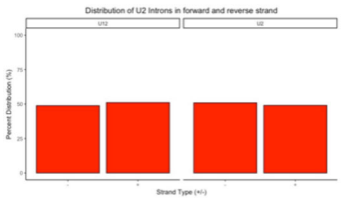

Figure 5. Distribution of minor introns in human DNA strand

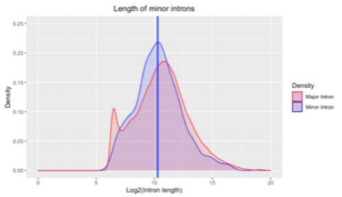

Figure 6. Distribution of minor intron length the blue line indicates the median of the minor intron length.

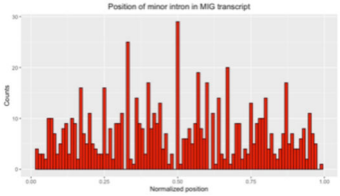

Figure 7. Distribution of minor intron position within mRNA transcript

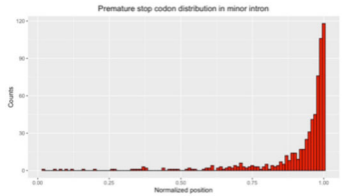

Figure 8. Distribution of PTC in minor intron

Reference

Burge CB, Padgett RA, Sharp PA. (1998). Evolutionary fates and origins of U12-type introns. Mol Cell 1998, 2:773–785. doi: 10.1016/S1097-2765(00)80292-0
Farach, L. S., Little, M. E., Duker, A. L., Logan, C. V., Jackson, A., Hecht, J. T. and Bober, M. (2018). The expanding phenotype of RNU4ATAC pathogenic variants to Lowry Wood syndrome. Am. J. Med. Genet. A 176, 465–469.
He, H., Liyanarachchi, S., Akagi, K., Nagy, R., Li, J., Dietrich, R. C., Li, W., Sebastian, N., Wen, B., Xin, B. et al. (2011). Mutations in U4atac snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. Science 332, 238-240.
Madan, V., Kanojia, D., Li, J., Okamoto, R., Sato-Otsubo, A., Kohlmann, A., Sanada, M., Grossmann, V., Sundaresan, J., Shiraishi, Y., Miyano, S., Thol, F., Ganser, A., Yang, H., Haferlach, T., Ogawa, S., & Koeffler, H. P. (2015). Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. Nature communications, 6, 6042. https://doi.org/10.1038/ncomms7042
Merico, D., Roifman, M., Braunschweig, U., Yuen, R. K. C., Alexandrova, R., Bates, A., Reid, B., Nalpathamkalam, T., Wang, Z., Thiruvahindrapuram, B. et al. (2015). Compound heterozygous mutations in the noncoding RNU4ATAC cause Roifman Syndrome by disrupting minor intron splicing. Nat. Commun. 6, 8718.
Nagy, E. and Maquat, L. E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem. Sci. 23, 198-199.
Olthof, A.M., Hyatt, K.C. & Kanadia, R.N. Minor intron splicing revisited: identification of new minor intron-containing genes and tissue-dependent retention and alternative splicing of minor introns. *BMC Genomics* **20**, 686 (2019). https://doi.org/10.1186/s12864-019-6046-x
Patel, A. A., McCarthy, M., & Steitz, J. A. (2002). The splicing of U12-type introns can be a rate-limiting step in gene expression. *The EMBO journal*, *21*(14), 3804–3815. doi:10.1093/emboj/cdf297
Tarn, W-Y., and Steitz J.A. (1996). Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. Science 273, 1824–1832. doi: 10.1126/science.273.5283.182
Turunen, J. J., Niemelä, E. H., Verma, B., & Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. *Wiley interdisciplinary reviews. RNA*, *4*(1), 61–76. doi:10.1002/wrna.114
Wahl MC, Will CL, Lührmann R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136:701–18. doi: 10.1016/j.cell.2009.02.009.
Younis I, Dittmar K, Wang W, Foley SW, Berg MG, Hu KY, Wei Z, Wan L, Dreyfuss G. (2013). Minor introns are embedded molecular switches regulated by highly unstable U6atac snRNA. *eLife*; 2:e00780. doi: 10.7554/eLife.00780

Result and Discussion

696 Minor introns were identified in our study. Among the 696 minor introns, 83% (640 minor introns) were embedded in protein coding genes and 7.27% (56 minor introns) were embedded in non-protein coding genes (Figure 1). The majority (92.58%) of minor introns have a single minor intron and remaining few have two or more (Figure 2). The research by Patel et al (2002) shows minor intron splicing inefficiency in human HeLa cell. With that in mind, we assumed that while MIGs with a single minor intron already experience difficulty in splicing, MIGs with more than one minor intron will experience stronger rigor in splicing. Thus, some regulatory mechanisms that upregulate the level of minor spliceosome are may be involved for multiple minor intron-containing genes (MMIGs). Of note, many MMIGs belong to the superfamily of voltage-gated ion channels. More studies in minor intron splicing regulation in MMIGs may provide important clues to determine the role of minor introns.
The minor intron subtype distribution result shows that 55.5% of minor introns are GTAG subtype and 11 other subtypes that are presumably recognized by minor spliceosome (Figure 3). In biological point of view, it is inefficient for minor spliceosome because large distribution of minor intron subtypes complicates the recognition of minor intron by minor spliceosome. Or, perhaps, having a range of minor intron subtype confers a regulatory function such that different subtypes have their own unique recognition processes and different splicing kinetics. This raises a possibility that the minor intron splicing is a selective process not a stochastic event.
In the statistical analysis, the minor intron distribution in human chromosome shows distribution bias at chromosome 19 and 9 – highest and lowest distribution respectively (Figure 4). This result tells us that the distribution of minor intron in chromosome is not random and that there could be some sort of reason or advantage of having minor intron in a specific chromosome. In other perspective, the position of minor intron within chromosome could be important. Under the presumption that the minor introns need to be preserved and transferred down the generation, the genes that harbor minor intron should be positioned in a region immune to mutations, transposon, DNA recombination, etc.
The minor intron length distribution analysis (avg. length = 4627 nt ; median = 1292 nt) shows that the minor introns are often "long" and would most likely to trigger alternative splicing (Figure 5). 27% of the minor introns have length less than 600 nt and the MIGs possessing these minor introns would mostly likely have a high retention rate. Since the intron retention has a possibility of trapping transcripts within the nucleus, one could presume that the MIGs with shorter minor introns provide rapid expression upon upregulation of minor intron splicing factors as the regulatory mechanism described in the research by Younis et al (2013) – stress-activated p38MAPK stabilizes the level of U6atac allowing an expression of hundreds minor intron-containing mRNA transcripts.
The intron retention gives two possible endings for transcripts retaining minor intron: 1) undergo degradation via NMD 2) gets translated to produce truncated protein. The first scenario will result in downregulation of the expression while the second scenario will result in the production of truncated proteins which may elicit some other effect. Thus, next task is to determine the necessary factors that dictate the fate of minor intron-containing transcripts. The study by Nagy et al (1998) shows that PTC positioned 50-55 nt upstream of the exon-exon junction triggers NMD. Thus, identifying the length of the ensuing exon (with respect to minor intron) will narrow down a list of MIGs that undergoes NMD.

Conclusion & Future Direction

In this research, we have statistically analyzed minor introns leveraging two publicly available data: MIDB and NCBI RefSeq. Our results provided conducive evidences in identifying the minor intron splicing regulation and paved a way for further analysis. The next task is to refine the minor intron search algorithm to determine the unidentified minor introns and factor their information into the statistical analysis. Also, with regards to the analysis, minor intron distribution in chromosome, one could normalize to the number of genes in the chromosome to find out different result and a potential link to the role of minor introns. Moreover, gathering experimental evidences of minor intron retention as well as production of truncated protein will be conducive to understand the molecular mechanisms of minor intron splicing regulation