Computational Analysis of Human mRNAs Regulated by GRSF-1, a Positive Regulator of Influenza Virus

Hikaru Taniguchi¹,² hikaru@sfc.keio.ac.jp
Kosuke Fujishima¹,² fujishima@sfc.keio.ac.jp
Yuka Watanabe¹,² yuka@sfc.keio.ac.jp
Yuko Osada¹ osada@sfc.keio.ac.jp
Masaru Tomita¹,²,³ mt@sfc.keio.ac.jp
Rintaro Saito¹,³ rsaito@sfc.keio.ac.jp
Akio Kanai¹,³ akio@sfc.keio.ac.jp

¹ Institute for Advanced Biosciences, Keio University, Tsuruoka, 997-0017, Japan
² Graduate School of Media and Governance, Keio University, Fujisawa, 252-8520, Japan
³ Department of Environmental Information, Keio University, Fujisawa, 252-8520, Japan

Keywords: 5’ UTR, translational control, influenza virus, GRSF-1

1 Introduction

Influenza viruses are well studied single strand RNA (ssRNA) viruses which infect human and several other vertebrates. Since translation of influenza viruses’ mRNAs are performed by hijacking host cell translation machinery, a mechanism of the translational up-regulation may also work in the host cell. Translation efficiency of influenza A virus mRNA is known to be very high, and this viral mRNA translational promotion is reported to require Guanine rich sequence factor 1 (GRSF-1), an RNA binding protein provided by host cell [1]. GRSF-1 binds to a sequential motif represented as 5’ – AGGGT – 3’ (or similar variant such as 5’ – AGGGGT – 3’), found in 5’ untranslated region (5’ UTR) of influenza A virus mRNA [1]. Based on the fact that this protein is coded in host genome, GRSF-1 mediated up-regulation may have been primary evolved in host cell. If so, GRSF-1 binding motif should be present in 5’ UTR of human mRNAs within some rule as a consequence of its translational regulation. Here, we performed computational analyses against 14506 full length cDNA data provided by H-invitational [2, 3] to find out possible mechanism of human mRNA regulation by GRSF-1. As a result, we revealed that the appearance of GRSF-1 binding motif was concentrated at a specific position of human mRNA 5’ UTR, 7 bp downstream of the transcription start site.

2 Method and Results

2.1 Frequency Analysis of GRSF-1 Binding Site

Number of GRSF-1 binding site observed in 5’ UTR of human mRNA and influenza A virus were compared. Forty mRNAs from the complete genome of five influenza A virus (Puerto Rico 1934, Hong Kong 1999, Korea 1968, New York 2004, Guangdong 1996) were obtained from NCBI Genome Database [4], and 14506 human full length cDNA that contained 5’ UTR sequence were obtained from H-invitational [2, 3]. To confirm the required condition of the motif position, firstly its frequency was calculated for each position from 5’ end and from start codon, and secondly the frequency was normalized by the number of 5’ UTRs available at each position. We found a peak of frequency in 5’ UTR of human cDNAs at the position 7 bp downstream of the 5’ end (Fig. 1A). Similar peak was observed in 5’ UTRs of influenza A virus at 10 bp downstream of the 5’ end (Fig. 1B), indicating that position from 5’ end of mRNAs is important for the GRSF-1 mediated translation activation.

2.2 Possible Functions of Human Genes with GRSF-1 Binding Site

From 14506 full length cDNA data, 56 cDNAs with 5’ – AGGGT – 3’ at the position 7 bp downstream of 5’ end were extracted and 16 cDNAs containing 5’ – AGGGGT – 3’ at the same position were also extracted. 25 out of the 72 were annotated as hypothetical protein or questionable transcript, and 44 possessed a conserved sequence 5’ – GGCACG – 3’ just before the motif, which was not observed in 5’ UTR of influenza A virus
mRNA. These 44 cDNAs are involved in many cellular processes such as transcription, sugar metabolism and development. Notably, the length of those 5' UTRs of human cDNA were mostly less than 60 bp (81.8% of 44), resembling 5' UTR of influenza A virus mRNA (100% of 40) despite the trend was not observed in the whole population (16.8% of 14506).

![Figure 1: Distribution of GRSF-1 Binding Site](image)

GRSF-1 Binding Site in 5' UTR of Influenza A Virus (A), Human full length cDNA (B). Horizontal axis represents position counted from 5' end of each mRNAs, while vertical axis represents normalized frequency of mRNAs which have GRSF-1 binding site at the position. Numbers in horizontal axis indicate nucleotide positions relative to the transcriptional start site.

3 Discussion

Genome wide analysis against 5' UTR of human full length cDNA was performed and GRSF-1 binding site was shown to emerge with distinctly high level at 7 bp distant from 5' end. Although experimental validation is necessary to confirm if they indeed interact with the regulatory protein GRSF-1, the sequential trend around the position where the motif frequently appeared suggests that the region is under pressure of some position specific selection. Besides, there were two differences between 5' UTRs of human and influenza A virus. One is that the distance between 5' end and the position where the motif was found in influenza A virus (10 bp downstream) was longer than that in human (7 bp downstream). In addition, the position of GRSF-1 binding motif in influenza A virus is actually located further downstream from the 5' end of mRNAs due to its unique transcriptional mechanism called “cap-stealing”, which utilizes a fragment of endogenous mRNA (10 bp to 12 bp) as a transcriptional primer [5]. The other is that over half of 5' UTRs in human cDNA dataset containing GRSF-1 motif at the peak position also possessed another conserved sequence not observed in 5' UTR of influenza A virus mRNA. Further analysis to find out the cause of different requirements among the motif in both species and following GRSF-1 mediated up-regulating system may have potential to provide de novo aide for flu pandemic.

References