DOI 10.18699/SBB-2022-507

Epigenetics behind SARS COV2

Basu A.

Department of Biochemistry, Gurudas College, Affiliated to University of Calcutta, Calcutta, India
basuanamikaami@gmailcom

Key words: Human endogenous retroviruses (HERV) sequences, distal enhancer sequence, hypokalemia, *KCNJ3* gene, SARS CoV2

Motivation and Aim: There is a 100 % sequence similarity between SARS CoV2 and human genome for seventy-one nucleotide sequence. In human genome this seventy-one-nucleotide sequence, is located from region (154,867,731–154,867,802) of chromosome 2 with no gap identity. This region is a part of HERV-H Chr_2q24LTR repeat region (154,867,204–157,867,802) with sequence ID: (MT992309.1). Human endogenous retroviruses (HERV) sequences, contain long terminal repeats regions (LTRs). These endogenous retroviruses (ERVs) can control the transcription of neighboring genes which are related with some diseases e.g., cancer etc. Long terminal repeats (LTRs) can function as either promoters or enhancers that may regulate adjacent human genes. In this study, the influence of this similar seventy-one nucleotide sequence, present in LTR, on neighboring human host gene expression, is analyzed.

Methods: From different databases of Human endogenous retroviruses (HERV) sequences, the detail characteristics of HERV-int elements, present in chromosome 2, have been searched. The Candidate Cis-regulatory Elements (cCREs) present in human genome for all cell types, have been identified for the specific ERV element. cCREs are the subset of representative DNase hypersensitive sites across ENCODE and Roadmap Epigenomics samples that are supported by either modifications (H3Kme3 and H3K27ac) or CTCF-binding data. From SCREEN database, the nearest protein-coding genes are investigated in the nearby location, for that chromosome region of cCRE. The data from RAMPAGE (RNA annotation and Mapping of Promoters for Analysis of Gene Expression) allows the annotation and the expression of gene at transcription start sites with nucleotide precision. The specific association of HERV/LTRs and transcription factors are frequently observed in human genome. Identification of TF binding sites (TFBSs) on the specific HERVH-int/LTR and DNase I hypersensitive sites (DNSs) for this specific chromosomal region are important for epigenetic control of genes within the chromosome. From JASPAR CORE database, several transcription factor binding sites are available in the regulatory region of genome. Chromosome folding data from in situ Hi-C and Micro-C XL experiments on H1-hESC (embryonic stem cells) and HFFc6 (foreskin fibroblasts) cell lines can be obtained from the 4D Nucleome Data Portal. With these data the number interactions are estimated between regions of genome. Finally, the effect of cis regulatory elements and control of gene expression on neighboring gene in SARS CoV2 patients has been elucidated.

Results: The search for HERV-int element, ERV type in human endogenous retrovirus database (HERVd) (http://herv.img.cas.cz) shows 1,337 entities. Among them, ERV element, ERV_2041653 (LTR7, HERVH-int) is present in Chr 2 (154,867,229–154,874,560) region. By searching from different databases viz, HERVd, ENCODE etc., the details of ERV_2041653 have been characterized. The presence of a candidate cis-regulatory element as distal enhancer sequence (154,865,721–154,866,057) has been recognized in this region. This enhancer sequence is present outside the 2kb of a transcription start site, so it is considered as distal enhancer sequence. The DNase activity and histone protein modification scores for this enhancer sequence are collected from SCREEN database. With the help of dbHERV-REs, the regulatory elements associated with HERV/LTR can be recognized.

Thus, TF binding sites (TFBSs) and DNase I hypersensitive sites, along with regulatory elements present in this retrotransposon are identified for LTR7. High scores for DNase activity and acetylation in histone, reveals that this distal enhancer sequence plays important role in epigenetic control of nearby gene expression in human genome. *KCNJ3* (potassium inwardly rectifying channel subfamily J member 3) is a protein coding gene is present in the nearby location (154,698,748–154,856,459) of this distal enhance sequence (ID E2042853). The gene product of this gene helps to transport K^+ inside the cell. In case of tested COVID risk variants, two genetic variants are detected within the gene sequence of KCNJ3 gene. Since hypokalemia is detected in SARS CoV2 infected patients, so there may be a correlation between occurrence of hypokalemia in COVID-19 patients and activation of *KCNJ3* gene during the viral infection. During transcription of KCNJ3 gene, epigenetic control of this gene expression occurs through acetylation in histone protein present in nucleosome complex. Thus the packing in this complex becomes loose. This results in facilitating more transcription factors binding in the promoter region of the gene. Moreover that, the presence of DNSs on this chromosome region helps to access DNase I enzyme, results in the remodeled state of DNA for TF binding. In the heatmap of chromatin folding data, a high score is obtained between KCNJ3 gene and distal enhancer region. This enhancer region is present in the downstream region of the gene sequence in chromosome. This high score between two regions suggests that they are probably located in the close proximity in chromosomal three-dimensional space. Both *KCNJ3* gene and enhancer sequence comes in close proximity by chromatin looping method as shown in Fig. 1.

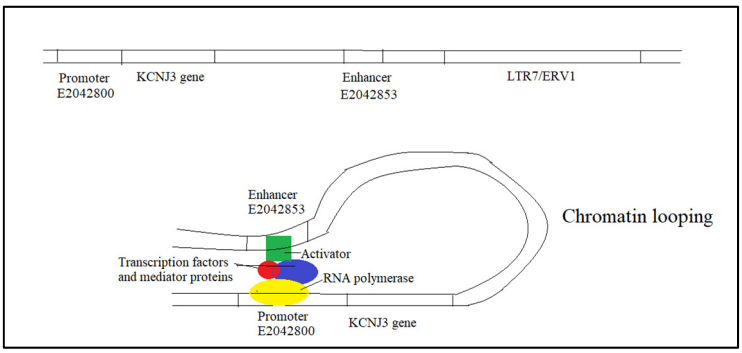

Fig. 1. Chromosome 2 remodeling causing hypokalemia in COVID-19 patients

In the distal enhancer region (E2042853), STAT6 has been selected, from JASPAR CORE 2022 database. Signal transducer and activator of transcription 6 (STAT6) can act as activator molecule during epigenetic control of *KCNJ3* gene expression due to presence of the downstream enhancer (E2042853). The target prediction study for transcription factor reveals that the *KCNJ3* gene is one of the target genes of the STAT6 transcription factor. The binding of activator protein STAT6, with enhancer sequence, facilitates the RNA polymerase binding in the promoter region of *KCNJ3* gene. Thus, the rate of gene expression of *KCNJ3* gene increases, which rises the K^+ concentration inside the cells, resulting in hypokalemia in blood serum of COVID-19 patients.

Conclusion: Due to presence of similar seventy-one nucleotide sequence, the incorporation of viral genome in human genome can occur. It facilitates the regulation of neighboring protein coding gene expression. A single nucleotide polymorphism (SNP) is found in the binding site of STAT6 TF. The altered allelic frequency from cytosine to thymine, causes inefficient TF binding in the promoter region of this KCNJ3 gene. This SNP in human genome, controls the gene expression of KCNJ3 gene and finally the frequency of occurrence of hypokalemia in COVID-19 patient, may be observed.