

Análise computacional do mecanismo de regulação do gene *LAMTOR5* em Humanos

Palavras-Chave: Bioinformática, Regulação de Transcrição, Hepatite-B

Autores(as):

Felipe dos Santos Minatogau – Ilum Escola de Ciência

João Pedro Aroucha de Brito – Ilum Escola de Ciência

Dr^a. Mariana Piccoli Gonçalves – LNBio, CNPEM

Prof^a. Dr^a. Juliana Helena Costa Smetana (orientador(a)), Ilum Escola de Ciência

INTRODUÇÃO:

O vírus da hepatite B (HBV) pertence à família hepadnaviridae é o causador da Hepatite B em humanos. Sabe-se que o seu material genético é um DNA circular parcialmente dupla fita. O genoma da dupla fita mede entre 3020 até 3320 nucleotídeos de comprimento, enquanto o genoma da fita parcial apresenta 1700 até 2800 nucleotídeos de comprimento. Entre os transcritos que são possíveis de serem gerados, um deles codifica a proteína X do vírus da hepatite B (HBx), uma proteína capaz de interferir com mecanismos regulatórios da célula hospedeira, sendo essencial para o sucesso da infecção. Um desses mecanismos parte da interação entre a HBx com a proteína *HBx Interacting Protein* (HBXIP), também conhecida como Lamtor5, uma subunidade do complexo Ragulator. A sua subunidade HBXIP é um elo entre a infecção viral do HBV e a maquinaria celular. Esse complexo está envolvido em diversas vias de sinalização sensíveis a disponibilidade de nutrientes. Dessa maneira, ele está relacionado com vias do tráfego lisossomal e na regulação da atividade metabólica celular (Zhang et al; Efeyan et al).

A proteína HBXIP é codificada por um único gene presente no cromossomo 1, chamado de *LAMTOR5*, e apresenta quatro isoformas diferentes de 79, 90, 91 ou 173 aminoácidos relacionadas com diferentes transcritos. As isoformas de 79 e 90 aminoácidos não são alvos desse estudo. A isoforma de 91 aminoácidos, também chamada de isoforma curta ou canônica, é expressa consistentemente e em todos os tecidos. A isoforma de 173 aminoácidos é conhecida como a isoforma longa, e é expressa restritamente e em baixo nível basal nos tecidos renais da medula e do córtex, no íleo do intestino delgado e no fígado. Com análises feitas *in silico* usando a *database* NCBI, descobriu-se que o mecanismo da geração da isoforma longa não parte de *splicing* alternativo, e sim de um diferente sítio de início de transcrição (SIT), anterior ao SIT da isoforma curta (Figura 1). Dessa maneira, dois diferentes transcritos são gerados, com diferentes códons de iniciação. O transcrito da isoforma longa apresenta um códon de iniciação com uma distância de 82 códons anterior ao códon de iniciação da isoforma curta, e portanto, a isoforma longa apresenta 82 aminoácidos a mais no N-terminal.

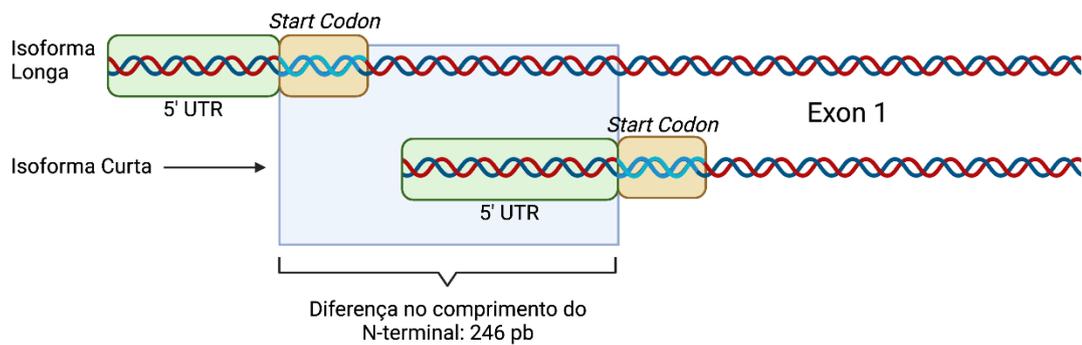


Figura 1: Diferença entre os transcritos da isoforma longa e isoforma curta de HBXIP. Imagem de autoria própria, feita no biorender.

Nesse projeto, estuda-se os mecanismos regulatórios da transcrição da isoforma longa da HBXIP. Dessa forma, entender o mecanismo pelo qual uma ou outra isoforma é expressa na célula em diferentes condições é um primeiro passo para entendermos a relação funcional de cada isoforma com a infecção viral. Além disso, serão apresentados os resultados dessas análises e possíveis estratégias para avançar com outras técnicas como edição gênica e silenciamento de RNA de maneira a entender os mecanismos de regulação da infecção viral por HBXIP.

METODOLOGIA:

Para fazer a análise dos elementos regulatórios, utiliza-se algumas ferramentas de análises de bancos de dados genômicos, destacando a base *UCSC Genome Browser* (Velmethed, D. et al.; Clawson, Hiram et al.) e o *NCBI Genome Data viewer* (Sayers, E. W. et al) Com essas ferramentas, é possível sobrepor diferentes *tracks*, com anotações e informações experimentais de diferentes técnicas. Para o projeto, é necessário descobrir dados sobre região promotora e o transcrito de RNA mensageiro da isoforma longa do gene *LAMTOR5*. As *tracks* usadas foram: *GENCODE V44* (Frankish, Adam et al.), um projeto de identificação e classificação de genes e suas estruturas; *FANTOM5* (Noguchi, S. et al.), um projeto com dados do transcriptoma como informações sobre os SIT de genes obtidos por CAGE-seq; e por último o *ENCODE* (*The Encode Project Consortium* et al.), um projeto com dados funcionais do genoma, como marcações em histonas relacionadas com elementos regulatórios. A Figura 2 mostra como são dispostos os dados presentes nos bancos genômicos e *tracks*, onde na esquerda

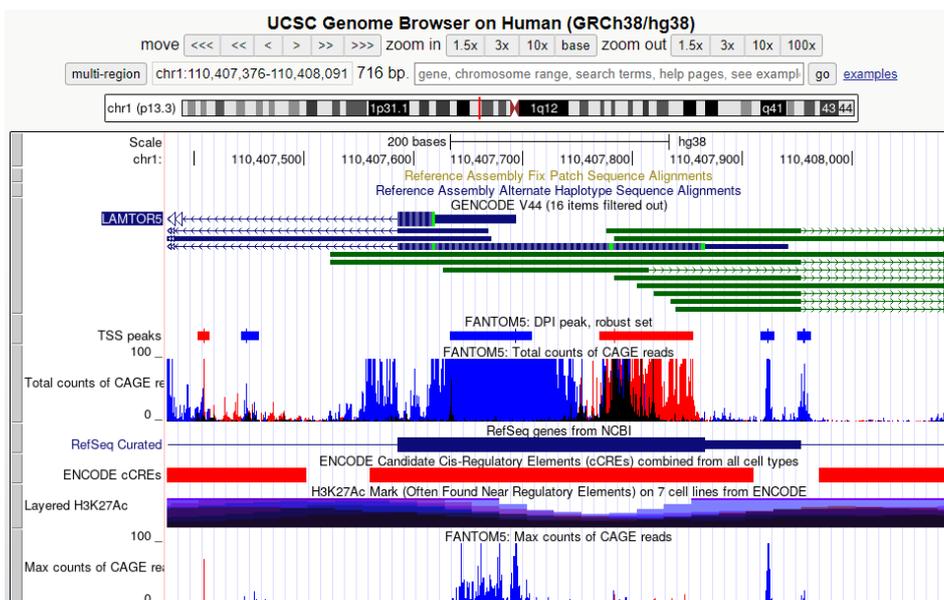


Figura 2: Captura de tela da interface do UCSC Genome Browser (Velmeshev et al., Clawson et al.). Pode-se ver a maneira como são expostos os dados da região genômica analisada.

RESULTADOS E DISCUSSÃO:

Foram obtidos como resultados os sítios de início de transcrição, tanto da isoforma longa e curta, e suas respectivas regiões promotoras. Destaca-se que foram encontrados na região promotora da isoforma longa duas possíveis *TATA-box* ineficientes, a primeira com 29 nt de distância do SIT, e a segunda com 44 nt de distância. Ambas não são regiões TATAA consenso, e por isso são ditas ineficientes, o que corroboraria com a baixa expressão basal da isoforma longa nos tecidos onde é expressa.

Dessa maneira, torna-se possível o desenho de estratégias que visam o nocaute da isoforma longa, sem que seja afetada a expressão da isoforma curta (Figura 3). Uma das técnicas que podemos propor é o *knockout* por CRISPR/Cas9, uma tecnologia que usa a enzima Cas9 para fazer um corte numa região altamente específica que é definida por uma fita de RNA guia. Se o corte ocorrer na sequência de *Kozak* do gene ou na sequência promotora, poderemos alterar respectivamente a tradução ou expressão da HBXIP longa, de maneira a diminuir total ou parcialmente a sua presença em uma linhagem celular.

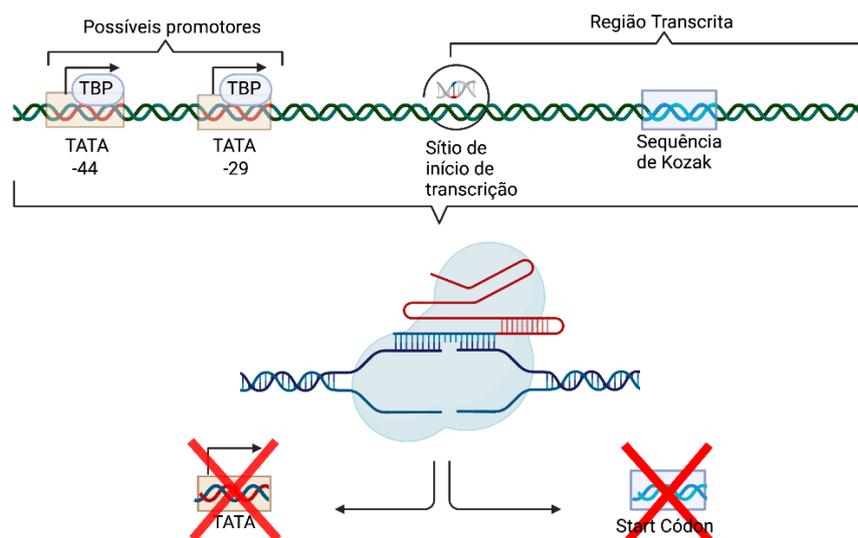


Figura 3: Sequência promotora e início da região de transcrição do gene *LAMTOR5*, isoforma longa. Essa região poderá servir de alvo para estratégias de knockout da isoforma longa da *HBXIP*, como o uso da enzima Cas9 associada a um RNA guia específico para a sequência de Kozak ou a TATABox promotora.

BIBLIOGRAFIA

- ZHANG, Tianlong et al. **Structural basis for Ragulator functioning as a scaffold in membrane-anchoring of Rag GTPases and mTORC1**. *Nature Communications*, v. 8, n. 1, p. 1394, nov. 2017. Disponível em: <<https://doi.org/10.1038/s41467-017-01567-4>>. DOI: 10.1038/s41467-017-01567-4.
- EFEYAN, Alejo et al. **Amino acids and mTORC1: from lysosomes to disease**. *Trends in Molecular Medicine*, v. 18, n. 9, p. 524-533, set. 2012. Disponível em: <https://doi.org/10.1016/j.molmed.2012.05.007>. DOI: 10.1016/j.molmed.2012.05.007.
- VELMESHEV, D. et al. **Single-cell analysis of prenatal and postnatal human cortical development**. *Science*, v. 382, n. 6667, p. eadf0834, 13 out. 2023. PMID: 37824647.
- CLAWSON, Hiram et al. **GenArk: towards a million UCSC genome browsers**. *Genome Biology*, v. 24, n. 1, p. 217, out. 2023. Disponível em: <https://doi.org/10.1186/s13059-023-03057-x>. DOI: 10.1186/s13059-023-03057-x.
- SAYERS, E. W. et al. **Database resources of the national center for biotechnology information**. *Nucleic Acids Research*, v. 50, n. D1, p. D20-D26, 7 jan. 2022. DOI: 10.1093/nar/gkab1112. PMID: 34850941. PMCID: PMC8728269.
- FRANKISH, Adam et al. **GENCODE 2021**. *Nucleic Acids Research*, v. 49, n. D1, p. D916-D923, dez. 2020. Disponível em: <https://doi.org/10.1093/nar/gkaa1087>. DOI: 10.1093/nar/gkaa1087.
- NOGUCHI, S. et al. **FANTOM5 CAGE profiles of human and mouse samples**. *Scientific Data*, v. 4, p. 170112, 2017. Disponível em: <https://doi.org/10.1038/sdata.2017.112>. DOI: 10.1038/sdata.2017.112.
- THE ENCODE PROJECT CONSORTIUM et al. **Expanded encyclopaedias of DNA elements in the human and mouse genomes**. *Nature*, v. 583, n. 7818, p. 699-710, jul. 2020. Disponível em: <https://doi.org/10.1038/s41586-020-2493-4>. DOI: 10.1038/s41586-020-2493-4.