

Desvendando o genoma escuro: o mistério além do DNA codificante

Giovanna Igepi^{1*}
 Hélio Gomes da Silva Filho²
 Natalia dos Santos²
 Fábio Porto-Foresti^{1,2,3}
 Ricardo Utsunomia^{1,2,3}

¹Programa de Pós-graduação em Ciências Biológicas (Genética), do Instituto de Biociências de Botucatu, Câmpus de Botucatu – UNESP.

²Programa de Pós-graduação em Biociências (Interunidades) da Faculdade de Ciências e Letras, Câmpus de Assis e da Faculdade de Ciências, Câmpus de Bauru – UNESP.

³Departamento de Ciências Biológicas, Faculdade de Ciências, Câmpus de Bauru – UNESP.

*g.igepe@unesp.br



ISSN 2237-8766

E-MAIL:
 APRENDENDO.CIENCIA@HOTMAIL.COM

Palavras-chave:

Biodiversidade

DNA

Genética

Genomas

Após 13 anos de estudos, o **sequenciamento** completo do **genoma** humano foi publicado, e em fevereiro de 2001, o acesso ao "livro da vida" foi permitido. O genoma humano contém informações para a codificação e formação de proteínas em mais de 200 tipos diferentes de células. No entanto, foi descoberto que menos de 2% das três bilhões de bases nitrogenadas do genoma humano estão ligadas à produção de proteínas, resultando na identificação de aproximadamente 20 mil **genes** codificadores. Esses números são curiosos, pois os geneticistas puderam comparar a quantidade de genes codificadores do corpo humano com os de vermes, que também possuem em média 20 mil genes, e moscas-das-frutas, com 13 mil. Essa descoberta foi um marco na ciência, uma vez que grande parte do **DNA** não está associada a genes codificadores. Mas, então, por que essas sequências estão presentes?

Inicialmente, essas sequências, cuja função era desconhecida, foram chamadas de "matéria escura". Com o tempo, os geneticistas renomearam esse conceito como "DNA lixo", acreditando que se tratava de restos de genes que um dia foram significativos, mas que perderam sua importância. No entanto, essa ideia foi contestada, e muitos defendem que há uma razão evolutiva para que o tamanho do genoma seja mantido.

Assim, 20 anos após a publicação do genoma humano, uma das funções do "DNA lixo" foi identificada: regular o processo de expressão dos genes produtores de proteínas, incluindo a produção de moléculas de **RNA** que participam da regulação das funções do organismo.

Uma das maiores dificuldades no estudo do genoma humano foi entender a funcionalidade das regiões não-codificadoras. Essas áreas contêm sequências repetidas de DNA conhecidas como **transposons**. Inicialmente, o foco do sequenciamento estava nas **regiões codificadoras** (exoma), mas com o tempo, tornou-se necessário estudar essas sequências repetidas para compreender sua origem, função e sobrevivência ao longo do tempo.

Em um experimento intitulado "Um **retrotransposon** específico de camundongos impulsiona uma **isoforma** conservada de *Cdk2ap1* essencial para o desenvolvimento", pesquisadores removeram um trecho específico de transposon em camundongos, resultando na morte de metade da prole antes do nascimento, indicando que essas sequências são importantes para a sobrevivência. Mas se

os transposons são sequências repetidas, de onde vêm?

Sugere-se que a existência dos transposons no genoma seja de origem viral e invasora, possivelmente derivada de patógenos que infectam células germinais, as células responsáveis por formar a próxima geração. Essas sequências possuem a propriedade de mobilidade, podendo transpor-se ao longo do genoma, daí o nome transposon (Figura 1). O movimento dos transposons pode provocar **mutações** gênicas, gerando curiosidade sobre suas consequências futuras.

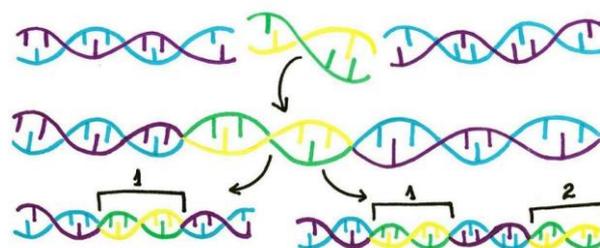


Figura 1. Representação esquemática do processo de inserção de transposons no genoma. À esquerda, observa-se a inserção inicial de um transposon em um sítio genômico. À direita, o transposon sofre duplicação e é inserido em um novo local do genoma, caracterizando um evento típico de retrotransposição. **Fonte:** elaborado por Giovanna Igepi.

Um estudo intitulado “A base genética da evolução da perda da cauda em humanos e primatas” sugere que uma inserção de um elemento transponível no genoma humano, levou à perda da cauda na linhagem dos primatas, o que favoreceu o desenvolvimento da locomoção ereta. Além de influenciar o desenvolvimento, a mobilidade das sequências repetidas também pode ser responsável por doenças. Estudos de associação genômica ampla (GWAS) identificaram doenças como diabetes, Alzheimer e cardiopatias localizadas em regiões do genoma onde não há **transcrição e codificação de proteínas**.

Com o avanço das pesquisas sobre doenças, ancestralidade e a funcionalidade do "DNA lixo", pesquisadores passaram a se referir a ele como "genoma escuro", um mistério a ser desvendado ao longo dos anos. Foi identificado que a inserção de um retrotransposon no íntron do gene TAF1 está ligada à distonia-parkinsonismo, uma condição observada frequentemente na população de Panay, nas Filipinas. A variante do gene TAF1, que apresenta uma inserção de retrotransposon, resulta em níveis reduzidos de expressão do gene, o que pode justificar a manifestação do parkinsonismo.

Além das doenças, o "genoma escuro" pode fornecer instruções para a formação de diversas moléculas, incluindo RNAs não-codificantes, que podem regular a atividade gênica. Esses RNAs são essenciais na leitura das informações do DNA pela **RNA polimerase** e na tradução das informações pelo **RNA ribossômico**, com a ajuda dos **RNAs transportadores**. A pesquisa sobre essas sequências repetidas ainda é preliminar, e muitas perguntas permanecem sem resposta. Por exemplo, por que essas sequências se conservam ao longo de milhares de anos? O genoma humano continua a evoluir? A ciência tem um papel crucial em estudar, experimentar e discutir os achados, e a busca por respostas é essencial para o progresso da humanidade.

GLOSSÁRIO

Codificação de proteína – após a transcrição do DNA, forma-se uma molécula de RNA mensageiro, composta por uma fita simples que é lida em segmentos de três bases nitrogenadas, chamados códon. Cada códon determina um aminoácido específico, que será incorporado para formar a proteína.

DNA (Ácido desoxirribonucleico) – macromolécula composta de nucleotídeos

que contém toda informação genética de um organismo. O DNA é formado por uma dupla fita, no formato de hélice, composta por 4 tipos de nucleotídeos, adenina, timina, citosina e guanina, as quais compõem os genes e servem de código de instrução para produção de proteínas.

Gene – unidade funcional da hereditariedade.

O gene apresenta sequências de nucleotídeos que contém regiões de transcrição para produção de proteínas.

Genoma – conjunto total de informação genética de um indivíduo presente em uma célula.

Isoforma – são variantes de um mesmo gene ou de genes diferentes.

Mutações – alteração no DNA em uma determinada região, como troca, duplicação ou deleção de bases nitrogenadas, podendo ocorrer por erros durante a replicação do DNA, exposição a radiações e/ou substâncias químicas mutagênicas.

Regiões codificadoras – conhecidas também como éxons, são regiões do DNA que apresentam sequências de nucleotídeos utilizados para sintetizar proteínas.

Retrotransposons – sequências móveis de DNA que tem o poder de se replicar ao longo do genoma.

RNA (Ácido ribonucleico) – molécula responsável pela síntese de proteínas, composta por uma fita simples, contendo fosfato, bases nitrogenadas (adenina, uracila, citosina e guanina) e o açúcar ribose.

RNA polimerase – enzima que realiza a transcrição da fita de DNA em RNA mensageiro.

RNA ribossômico (RNAr) – estrutura encontrada no citoplasma celular e no retículo endoplasmático rugoso, constituindo o ribossomo, responsável pela síntese de proteínas.

RNA transportador (RNAt) – realiza o transporte do aminoácido específico até o ribossomo, que contém um anticódon que se liga ao códon do RNA mensageiro, formando uma cadeia de aminoácidos para síntese proteica.

Sequenciamento – técnica utilizada para identificar a sequência dos nucleotídeos, uma leitura do genoma.

Transcrição – processo de síntese de RNA a partir da leitura das bases nitrogenadas da fita de DNA, formando uma fita unifilamentar de RNA mensageiro complementar a fita de DNA, com a troca da base nitrogenada timina por uracila no RNA.

Transposon – sequência móvel de DNA que

tem o poder de se deslocar e se transpor para diferentes regiões do genoma.

REFERÊNCIAS BIBLIOGRÁFICAS

Camargos, S.T. 2008. **Caracterização fenotípica e genotípica de parkinsonismo e distonia familiares no ambulatório de distúrbios de movimento do Hospital das Clínicas da Universidade Federal de Minas Gerais**. Tese (Doutorado em Neurologia) - Faculdade de Medicina, UFMG, Belo Horizonte. Disponível em: <http://hdl.handle.net/1843/ECJS-7GJGLB>. Acesso em: 14 out. 2024.

Góes, A.C.S.; Oliveira, B.V.X. 2014. Projeto Genoma Humano: um retrato da construção do conhecimento científico sob a ótica da revista ciência hoje. **Ciência e Educação**, 20: 561-577. DOI: <http://dx.doi.org/10.1590/1516-73132014000300004>

Modzelewski, A.J.; Shao, W.; Chen, J.; Lee, A.; Qi, X.; Noon, M.; Tjokro, K.; Sales, G.; Biton, A.; Anand, A. A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. **Cell**, 184: 5541-5558. DOI: <http://dx.doi.org/10.1016/j.cell.2021.09.021>

Xia, B.; Zhang, W.; Zhao, G.; Zhang, X.; Bai, J.; Brosh, R.; Wudzinska, A.; Huang, E.; Ashe, H.; Ellis, G.; Pour, M.; Zhao, Y.; Coelho, C.; Zhu, Y.; Miller, A.; Dasen, J.S.; Maurano, M.T.; Kim, S.Y.; Boeke, J.D.; Yanai, I. 2024. On the genetic basis of tail-loss evolution in humans and apes. **Nature**, 626, 1042-1048. DOI: <https://doi.org/10.1038/s41586-024-07095-8>