

Avanços da genómica

e os desafios na descodificação da vida

Agostinho Antunes
CIIMAR/FCUP

CITAÇÃO

Antunes, A. (2022)
Avanços da genómica,
Rev. Ciência Elem., V10(04):056.
doi.org/10.24927/rce2022.056

EDITOR

João Nuno Tavares
Universidade do Porto

EDITOR CONVIDADO

Alexandre Lopes Magalhães
Universidade do Porto

RECEBIDO EM

05 de dezembro de 2022

ACEITE EM

05 de dezembro de 2022

PUBLICADO EM

20 de dezembro de 2022

COPYRIGHT

© Casa das Ciências 2022.
Este artigo é de acesso livre,
distribuído sob licença Creative
Commons com a designação
[CC-BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/), que permite
a utilização e a partilha para fins
não comerciais, desde que citado
o autor e a fonte original do artigo.

rce.casadasciencias.org



A “conclusão” em 2003 do Projeto Genoma Humano proporcionou um avanço sem precedentes no estudo da vida. A compreensão do DNA, incluindo o recurso a manipulações genéticas, abriu as portas às edições de genomas, à experimentação de mutações do DNA (relação genótipo-fenótipo), decifrando-se a base genética de várias doenças, alterações epigenéticas, elucidando possíveis tratamentos genéticos, criação de organismos capazes de produzir remédios, ou produtos industriais de base orgânica, etc. O progresso na sequenciação de genomas permite agora a geração em larga escala de genomas de referência e ambiciona-se assim representar a biodiversidade global, o que irá certamente revolucionar a genómica da conservação. Volvidos cerca de 20 anos da publicação do primeiro rascunho da sequência de todo o genoma humano, os investigadores mergulharam ansiosamente na era “pós-genómica”, remodelando a investigação biológica e abrindo portas para a medicina personalizada.

Os princípios básicos da genética foram lançados pela primeira vez por Gregor Mendel, em 1866, um monge da Ordem de Santo Agostinho que realizou experiências de cruzamentos controlados em ervilhas. Mendel esclareceu a forma como as características (por exemplo: cor, forma) são transmitidas de geração em geração e introduziu os termos “recessivo” e “dominante” para explicar a relação desses traços na descendência. Descreveu a ação de fatores “invisíveis” no fornecimento de traços visíveis de maneiras previsíveis (esses traços “invisíveis”, correspondem na realidade aos genes).

Mendel é amplamente considerado o pai da genética, mas essa notoriedade apenas foi reconhecida muito posteriormente. Bastante à frente do pensamento da época, as suas investigações foram apenas consideradas três décadas depois, em 1900.

O botânico e geneticista holandês Hugo de Vries, o botânico e geneticista alemão Carl Erich Correns e o botânico austríaco Erich Tschermak von Seysenegg, redescobrem independentemente o trabalho de Mendel, obtendo resultados de experiências de hibridização semelhantes. Na Grã-Bretanha, o biólogo William Bateson, tornou-se um dos principais defensores das teorias de Mendel e reuniu vários entusiastas conhecidos como “mendelianos”. Inicialmente, esses defensores entraram em conflito com os “darwinistas” (apoiantes da teoria de evolução de Charles Darwin, que na altura defendiam que a evolução se baseava na seleção de pequenas variações que se misturavam, mas as variações de Mendel claramente não se misturavam). A conciliação entre a teoria mendeliana e a teoria evolutiva aconteceria apenas passadas três décadas.

A compreensão dos princípios da herança genética foi avançando consideravelmente, e os genes passaram a ser considerados as unidades discretas da hereditariedade (gerando as enzimas controladoras das funções metabólicas), mas apenas em 1944, Oswald Avery, imuniquímico do Hospital do Rockefeller Institute for Medical Research, identifica o ácido desoxirribonucleico (DNA) como o “princípio transformador”. Na bactéria responsável pela pneumonia, o pneumococo, se a uma forma viva, mas inofensiva, de pneumococo fosse misturada uma forma inerte, mas letal, a bactéria inofensiva tornar-se-ia mortal — a substância não parecia ser uma proteína ou carboidrato, mas sim um ácido nucleico e subsequentes análises mais aprofundadas, revelaram ser o DNA.

O trabalho de Avery e dos seus colaboradores sobre o DNA lançou as bases para uma das maiores descobertas do século XX. Em 1953, James Watson e Francis Crick, com base em dados de raios-X disponíveis e na construção de modelos, decifram a estrutura de dupla hélice do DNA (Prémio Nobel em 1962).

Como surgiu a revolução no conhecimento genómico?

A descodificação do DNA só viria a ser delineada com o desenvolvimento de técnicas rápidas de sequenciamento de DNA, desenvolvidas em 1977 por Frederick Sanger (Prémio Nobel em 1980). Esse avanço tecnológico iria possibilitar um desafio pioneiro — o Projeto Genoma Humano — oficialmente iniciado em 1990 com o Departamento de Energia (DOE) e os Institutos Nacionais de Saúde (NIH), nos EUA, e com previsão de 15 anos. Os objetivos incluíam: mapear o genoma humano e determinar as suas 3,2 giga bases (nucleótidos), bem como mapear e sequenciar os genomas de outros organismos úteis para o estudo da biologia, desenvolver tecnologia para a análise de DNA e estudar as implicações sociais, éticas e legais da investigação do genoma.

Subsequentemente, começaram a ser sequenciados os primeiros genomas de seres vivos. Em 1995, com base na nova estratégia de sequenciação “*shotgun*”, J. Craig Venter e colegas sequenciam o primeiro genoma de uma bactéria auto-replicante e de vida livre, o *Haemophilus Influenzae*, causador de meningite e de infeções respiratórias e do ouvido em crianças. No ano 2000, é descodificado o genoma da mosca-da-fruta (*Drosophyla melanogaster*) com 13 601 genes, o organismo mais complexo descodificado na época (as células humanas contêm 70 000 genes, mas o genoma humano ainda tinha um longo percurso a percorrer). Em 2002, é sequenciado o genoma do primeiro mamífero — o ratinho *Mus musculus*, permitindo comparar, pela primeira vez, o genoma humano em preparação com o de outro mamífero, revelando que ambos os genomas possuíam cerca de 90% de regiões alinháveis homólogas.

Em 2003, o Projeto Genoma Humano foi finalmente “concluído”, um projeto notável de esforços conjuntos do *International Human Genome Project (HGP) Consortium* e o *Celebra Genomics*, numa cerimónia apresentada na Casa Branca, com repercussões notáveis e equiparáveis ao feito de colocar o Homem na Lua. Em menos de 150 anos, desde os pioneiros como o Mendel, que demonstrou que características poderiam ser herdadas, é impressionante o avanço científico alcançado.

Esta forma final do genoma humano contém 2,85 giga bases (nucleótidos) e uma taxa de erro de apenas 1 evento em cada 100 000 bases sequenciadas (em 2001, o Projeto Genoma Humano publicou um “rascunho” cobrindo 90% do genoma humano, e subsequentemente foram preenchidas as lacunas em regiões ambíguas do DNA até se completar

99% do genoma humano). Surpreendentemente, foi identificado um número relativamente pequeno de genes que codificam proteínas (entre 20 000 e 25 000), havendo genes semelhantes com as mesmas funções presentes em diferentes espécies.

A sequenciação do genoma humano levantou também uma importante questão biológica: decifrar corretamente o genoma de um organismo requer a sua comparação com os genomas de outros organismos, partilhando ancestrais evolutivos comuns a diferentes escalas temporais. A genômica comparativa surge assim como um ramo de estudo prolífico, gerando-se imensos avanços científicos: percentagem de genes partilhados no Homem relativamente a diferentes espécies, por exemplo: chimpanzé (98%), ratinho (92%), mosca-da-fruta (44%), levedura (26%), planta (18%). Várias propostas são elaboradas para a sequenciação do genoma de diversos organismos, com base na relevância biológica, económica e saúde humana. São organizados vários consórcios internacionais para fomentar esse avanço científico, por exemplo: *Genome10K*, *i5K*, *Global Invertebrate Genomics Alliance*, *Fish10K*, *Bat10K*, *Vertebrate Genome Project*, *Earth BioGenome project*, *European Reference Genome Atlas*, etc.).

Os cientistas de todo o mundo continuaram a desenvolver a sua compreensão do DNA, incluindo o recurso a manipulações genéticas, criando com sucesso um organismo com um código genético artificial expandido, abrindo portas às edições de genomas. Isto permitiu a experimentação de mutações do DNA (relação genótipo-fenótipo), o tratamento de algumas doenças genéticas, a criação de organismos capazes de produzir remédios, ou produtos industriais de base orgânica.

Simultaneamente, vários avanços na área médica são alcançados, decifrando-se a base genética de várias doenças, com centenas de genes envolvidos, incluindo distúrbios complexos como a esquizofrenia, e impulsionando a produção de novos medicamentos.

Qual o estado da arte da compreensão do nosso genoma?

Desde o seu lançamento, o genoma humano de referência cobriu apenas a fração eucromática do genoma, deixando importantes regiões heterocromáticas inacabadas. Produzir uma sequência completa e sem lacunas do genoma humano foi uma meta para os biólogos, um sonho que se realizou em 2022.

A sequenciação de leitura longa (*long reads*) apoiou a montagem completa de telómero a telómero (T2T) da linhagem de células humanas pseudo-haplóides CHM13. Abordando os 8% restantes do genoma, o Consórcio Telomere-to-Telomere (T2T) apresenta a sequência completa de 3,055 giga bases de um genoma humano, T2T-CHM13, que inclui montagens sem lacunas para todos os cromossomas, exceto o Y, corrigindo erros anteriores e apresentando quase 200 milhões de pares de bases de sequências contendo 1 956 previsões de genes, 99 dos quais são previstos como codificadores de proteínas. As regiões completas incluem todas as matrizes de satélites centroméricos, duplicações segmentares recentes e os braços curtos de todos os cinco cromossomas acrocêntricos, revelando essas regiões complexas do genoma para estudos funcionais e variação genética.

As vastas melhorias na descoberta de variantes em amostras de diversos ancestrais posicionam o genoma de referência preciso e completo para revelar a diversidade entre as populações humanas. Adicionalmente, revelam a importância de aumentar os esforços para alcançar também montagens de nível T2T para primatas não humanos e outras espécies. Essa informação ajudaria a entender completamente a complexidade e o impacto das inova-

ções genômicas observadas, anteriormente desconhecidas para algumas das regiões com evolução mais dinâmica no nosso genoma.

São alavancadas inovações em tecnologia, *design* de estudo e parcerias globais com o objetivo de construir a referência de pangenoma humano da mais alta qualidade possível (*Human Pangenome Reference Consortium*). Este avanço tecnológico irá permitir, futuramente, a montagem rotineira de genomas diploides completos em outras espécies. Com atenção às estruturas éticas, a referência do pangenoma humano conterà uma representação mais precisa e diversificada da variação genômica global, melhorará os estudos de associação gene-doença entre as populações, expandirá a investigação genômica para as regiões mais repetitivas e polimórficas do genoma, fomentando um recurso genético definitivo para o futuro da investigação biomédica e medicina de precisão.

O que o futuro reserva para a nossa compreensão da genômica?

A publicação dos primeiros rascunhos do genoma humano lançou uma nova era na descoberta biológica, reduzindo o número de genes esperados, mas expandindo enormemente a compreensão da regulação genética. Atualmente, são produzidos genomas individuais e análises genômicas às dezenas de milhares, mas persistem alguns dos mesmos conflitos centrais em relação à disponibilidade de dados, equidade e privacidade.

Volvidos cerca de 20 anos da publicação do primeiro rascunho da sequência de todo o genoma humano, os investigadores mergulharam ansiosamente na era “pós-genômica”, remodelando a investigação biológica e médica, assim como os desafios de recolha, curadoria e acesso de dados gerados pelos atuais projetos. Avanços das metodologias de sequenciação (notavelmente o uso de *long reads*) permitiram explorar, mais recentemente, genomas de grandes dimensões relativamente ao genoma humano (3 giga bases), como o genoma do grande tubarão branco (4,63 giga bases), o genoma do axolote (30 giga bases) ou o genoma do peixe pulmonado Africano (40 giga bases), e ainda a sequenciação de genomas de várias espécies extintas, como o periquito-da-Carolina ou o gato-de-cimitarra.

O progresso na sequenciação de genomas permite agora a geração em larga escala de genomas de referência. Ainda que a sequenciação de um genoma possa ser efetuada atualmente em poucos minutos, a sua montagem, anotação e decifração pode necessitar de inúmeros recursos humanos especializados e requerer um considerável tempo (FIGURA 1).



FIGURA 1. A montagem, anotação e decifração de um genoma pode requerer inúmeros recursos humanos especializados e um considerável tempo.

O mundo pós-genoma esperado para os próximos anos requer o desenvolvimento e implementação de mais ferramentas bioinformáticas e o recurso a metodologias de inteligência artificial.

Consideráveis progressos têm sido ainda obtidos no campo inovador da epigenética (o estudo das mudanças nos organismos causadas pela alteração da expressão genética). Um elevado número de mecanismos moleculares pode afetar a atividade dos nossos genes. Surpreendentemente, as nossas experiências e escolhas de vida podem também alterar a atividade desses mecanismos, alterando a expressão génica. Ainda mais fascinante, é que essas modificações na expressão génica podem ser herdadas, o que significa que as experiências de vida dos nossos ancestrais podem influenciar fundamentalmente a nossa constituição biológica. Estas descobertas terão um impacto dramático no futuro do sistema de saúde, realçando a importância das escolhas de estilo de vida, o que trará impacto nas gerações futuras.

Outro grande desenvolvimento futuro espectável associado à investigação genómica é o desenvolvimento da medicina personalizada com o uso de medicamentos precisos, para eliminar os efeitos de muitas doenças genéticas causadas por genes mutantes, que diferem de uma pessoa para outra. Ao identificar essas combinações, através da sequenciação do genoma do indivíduo, os medicamentos podem ser adaptados de um modo preciso, proporcionando o melhor tratamento possível.