

## O Genoma Humano: Sequências e Consequências

Foi percorrido um longo caminho até aos dias de hoje, desde que em 1953 Watson e Crick descobriram o modo como as 4 letras do alfabeto humano, A, C, G, e T (Adenina, Citosina, Guanina e Timina) do DNA se organizavam, em dupla hélice. Assiste-se neste momento a uma plétora de genomas que têm sido decodificados, entre os quais o do genoma humano. Para se chegar a este ponto foi necessário possuímos todas as técnicas de DNA recombinante. Desde a descoberta do processo de sequenciar, da qual foram pioneiros Maxam e Gilbert e posteriormente Sanger, até à automatização de todo este processo, permitiu que a sequenciação atingisse o ritmo actual.

O DNA humano sequenciado foi obtido a partir de um conjunto de 21 dadores que aceitaram fazer parte do grupo, de acordo com regras criteriosamente estabelecidas. Estes dadores correspondiam a 5 indivíduos, 2 do sexo masculino e 3 do sexo feminino, de diferentes étnias, nomeadamente de um indivíduo afro-americano, asiático-americano, hispano-mexicano e dois caucasianos.

O genoma humano possui os seus 2,91 Giga pares de bases distribuídos por 23 pares de cromossomas possuindo informação para aproximadamente 50 000 genes. É possível que, após uma análise mais detalhada o número de genes seja ligeiramente maior. Da análise dos genomas já conhecidos e disponíveis em banco de dados verifica-se que os seus tamanhos não são correlacionados com o número de genes. Curiosamente o genoma humano não é o maior de todos os sistemas biológicos pois que a *Amoeba dubia*, *Amoeba proteus* e o genoma do sapo possuem tamanhos respectivamente de 230, 100 e duas vezes maiores que o do genoma humano. Este facto levanta uma questão curiosa sobre a razão de sistemas biológicos tão pouco diferenciados como as amibas, conterem um genoma tão grande. Quase se poderá imaginar que as amibas serão repositórios de DNA mas com que finalidade?

Uma casualidade do processo da evolução resulta tanto do tamanho do genoma (raras vezes reflecte a complexidade do organismo) como do tamanho do cromossoma (raramente reflecte a importância ou número de genes que ele transporta). Para ilustrar este ponto, temos alguns exemplos. O cromossoma 2, o maior de todos os cromossomas, possui apenas 33 genes que aparecem duplicados no cromossoma 14. Como excepção o cromossoma Y, que sendo o mais pequeno, é o que contém também menos genes. O cromossoma 19 é de todos os cromossomas o que contém mais genes, cerca de 23 genes/Mb enquanto que, por exemplo o cromossoma Y e 13 contém apenas 5 genes/Mb. A percentagem do genoma, que é pois ocupada por genes corresponde apenas a 1,1 - 1,4 % e a percentagem de genes com função ainda não conhecida é cerca de 42 %, um valor que curiosamente se aproxima dos valores encontrados para outros sistemas biológicos, tais como a bactéria, a levedura, a planta *Arabidopsis thaliana* etc. Também facto curioso é o que diz respeito à pequena taxa de variação existente entre indivíduos (1pb/em 1 250 pb ou seja cada indivíduo apresenta uma semelhança de 99.9%) o que uma vez mais indica que as diferenças fenotípicas dependem da expressão das diferentes proteínas.

A sequenciação do genoma humano permitiu também mostrar que tal como no passado muitos dos dogmas da chamada Biologia Molecular tiveram que ser postos de parte, também o paradigma de que um gene corresponde apenas a uma proteína teve que ser revisto pois que a um gene podem corresponder várias proteínas. Foi pois, o conhecimento do genoma humano que mostrou que os mecanismos que estão na origem da produção de várias proteínas a partir de um mesmo gene é mais geral do que inicialmente se tinha pensado.

Cerca de 75% do DNA é ocupado pelo designado DNA egoísta (selfish DNA) ou seja, regiões que são possivelmente implicadas na regulação da expressão dos genes mas também ainda em mecanismos até hoje ainda não identificados. Nestes 75% reside um grande mistério, que irá ainda, sem dúvida, ser objecto de aturadas investigações.

Uma grande percentagem de pseudogenes foi detectada no genoma humano prevendo-se para o genoma inteiro cerca de 20 000. Localizam-se junto das regiões centroméricas, ao contrário dos da levedura que se encontram nas regiões teloméricas. Como o próprio nome indica os pseudogenes são genes que por variados mecanismos perderam a sua função, sendo considerados "fósseis moleculares".

Certas regiões dos genomas são também repetidas. Assim, por exemplo, a levedura *Saccharomyces cerevisiae*, conhecida pelo fermento do padeiro, possui duas vezes o seu genoma. Obviamente que se se observa uma tal duplicação na levedura, essa duplicação é muito maior no genoma humano. Por exemplo, o maior segmento duplicado corresponde a um grande bloco de 33 proteínas que existe tanto no cromossoma 14 como no cromossoma 2. Também entre o cromossoma 18 e o 20 aparecem duplicados 12 dos 64 pares de genes existentes nestes cromossomas. Estes genes homólogos do mesmo sistema biológico são designados por parálogos sendo possível que tenham ocorrido por pressões selectivas durante a evolução, por alterações do meio ambiente (ou outras) ou que possa ser também um mecanismo extra para permitir a sobrevivência ao sistema biológico.

As funções moleculares potenciais, atribuídas a 26 383 genes humanos distribuem-se entre genes que codificam as proteínas do metabolismo, do ciclo celular, dos transportadores, factores de transcrição, ou seja, as proteínas que desempenham as funções essenciais do ser vivo. Foi interessante verificar que a comparação entre genomas de vertebrados e invertebrados tomando como referência o Homem, a minhoca, e a mosca *Drosophila melanogaster* revela um conjunto de 1523 genes comuns aos 3 sistemas biológicos. Estes genes ortólogos (homólogos de sistemas biológicos

diferentes) estão envolvidos nas mesmas categorias funcionais sendo suficientes para assegurar as funções vitais. O grande desafio que se coloca à nova compreensão da Biologia Humana é o de prever a função, bem como desenhar novas estruturas moleculares, como um potencial para novos diagnósticos ou alvos para os novos produtos que venham a ser descobertos. Com o conhecimento dos muitos genomas passa-se da chamada era pre-proteómica (proteoma é o conjunto das proteínas expressas numa célula) em que as atribuições funcionais derivavam de hipóteses, que por sua vez conduziam a experiências que eram desenhadas para compreender processos celulares específicos, onde apenas uma proteína estaria implicada. Actualmente assiste-se à era proteómica em que novas ferramentas são utilizadas na análise global das proteínas sob um ponto de vista funcional. Cria-se um novo modo de se fazer investigação celular, o que conduzirá à determinação de estruturas tridimensionais e à descrição de redes de interacção das proteínas em que os processos tradicionais não poderão ser postos de parte e em que certamente a Bioquímica desempenhará um papel muito relevante.

É razoável prever que durante o século XXI muitas das doenças que afligem a humanidade - em particular as doenças crónicas que são uma espécie de praga do mundo desenvolvido - virão a ser controladas. A lista é longa encontrando-se entre estas as doenças neurodegenerativas (Alzheimer e Parkinson por exemplo), do sistema imune (artrites), cancro, e doenças do foro psiquiátrico (esquizofrenia), entre muitas outras.

É evidente que as doenças complexas são resultados da interacção de muitos genes e do ambiente. Será necessário saber como os genes são regulados, como interagem com outros genes, bem como conhecer os factores ambientais que influenciam a sua expressão.

São tantas as repercussões que o conhecimento do genomas fornecem que por limitação de espaço apenas referimos algumas "*consequências das sequências*". Oportunamente voltaremos a este assunto.