

Fundamentos, Métodos e Significado do Mapeamento do Genoma Humano

Human Genome

Carlos M. Manuel*

* Doutorado em Medicina; Especialista do Serviço de Anatomia Patológica do HAB, Professor Titular de Patologia da Faculdade de Medicina da Universidade Agostinho Neto- Luanda- Angola e Professor Convidado da Universidade de Lisboa e de Humsoldt/Berlim

Rev CSE 2007;1:28-33

INTRODUÇÃO

Abstraindo-nos da medicina helénica da era hipocrática, estruturada nesse tempo em torno dos humores do organismo humano, o ulterior desenvolvimento da ciência médica evoluiu em paradigmas sucessivos, designadamente (i) o referente ao desenvolvimento da anatomia normal desde Galeno a André Vesalius, (ii) o da fisiologia normal do organismo com William Harvey, (iii) o da anatomia patológica de Johannes B. Morgagni, (iv) o da anatomia microscópica com Marcello Malpighi, (v) o da histopatologia com Carl Ludwig Rudolf Virchow, e finalmente (vi) o molecular, traduzido no conhecimento da genética e da bioquímica morfofuncional celulares e dos fluidos, esta última sendo no fundo a reactivação do conceito de patologia humoral, no seu tempo bastante criticada, da autoria do vienense Carl Friedrich von Rokitanski.

O conhecimento da estrutura e respectiva função endógena e de relação, em termos normais e anormais, constituiu sempre a base estruturante da ciência médica, nos povos do hemisfério ocidental com elementos culturais cujas origens remontam das civilizações helénica e romana.

É certo que ao binómio estrutura e função juntou-se na racionalização da génese dos processos patológicos o factor etiológico, podendo este ser endógeno ou exógeno ao organismo e, nesta última categoria, herdamos e desenvolvemos os trabalhos pioneiros de Louis Pasteur, Alexander Fleming, Charles Finlay e outros no domínio da microbiologia e infecciologia, Percival

Potter no domínio da cancerologia química, etc.

No que concerne aos factores etiológicos endógenos, que se traduzem em alterações dos elementos reguladores ou estruturantes do metabolismo (proteínas, hidratos de carbono e lípidos simples ou complexos), os respectivos processos metabólicos e patológicos que os produzem, são determinados pelos genes normais ou anormais que lhe estão subjacentes.

James Watson e Francis Crick elucidaram a estrutura do ADN, a macromolécula que integra os genes, constituindo a sua exacta localização nessa molécula, bases nucleotídicas que os compõem e função para que são responsáveis, o enorme desafio científico da actualidade, que elevará a ciência médica a um paradigma novo, superior e passível de reverter muito em medicina, que até aqui tem tido quase valor dogmático.

O presente artigo de revisão pretende actualizar o assunto, para leitores da comunidade médica e de biociências em geral, em conformidade com a sugestão do título acima enunciado.

EVOLUÇÃO DO CONHECIMENTO SOBRE GENÉTICA.

O início dos estudos sobre genética clássica é atribuído ao monge agostiniano Gregor Johann Mendel, cujos resultados dos trabalhos pioneiros de fertilização da planta endogâmica de jardim *Prisum sativum*, foram reunidos e apresentados em 1865 sob o título "Experiências sobre plantas Híbridas", no

Relatório da Sociedade de História Natural de Bruenn/Áustria. Desses trabalhos resultou a formulação das clássicas leis de hereditariedade resultantes da expressão de genes autossómicos segregados e não recombinantes [1,2].

Quatro anos após a apresentação dos resultados de Mendel, Friedrich Miescher viria a descobrir em 1869 o ADN, embora Miescher na altura lhe tivesse atribuído a designação de nucleína (em virtude de encontrar-se no núcleo celular).

Richard Altmann introduziu em 1889 o conceito de ácido nucléico para designar o ADN e sucessivamente até ao fim da primeira metade do século passado, Theodor Avery, Fred Griffith, Michael McLeod e outros [3], decifram a natureza química do DNA e MA, tendo Daniel Hershey e Michael Chase elucidado em 1952, o ADN como sendo a molécula portadora da informação genética, com base na marcação com fósforo e enxofre radioactivos, respectivamente do ADN e de proteínas parietais de bacteriófagos.

Finalmente em 1953, Francis Crick e James Watson elucidaram a estrutura definitiva em dupla hélix da molécula do ADN, tendo constituído essa descoberta o início da genética moderna e molecular, que tem permitido estudar melhor as unidades estruturais e funcionais constitutivas do ADN e MA, designadamente, os genes, os nucleótidos, as bases, os respectivos açúcares e grupos fosfatos.

Estrutura e Cinética da Transcrição do ADN

A molécula do ADN é constituída por duas “fitas” (que seja permitida a analogia apenas por razões didácticas) que se enrolam entre si em forma de hélix em torno de um eixo virtual. Cada “fita” é constituída por inúmeros nucleótidos, o que quer dizer, que cada uma que se enrola em espiral com a outra complementar, é uma estrutura polinucleotídica. Um gene corresponde a um segmento da dupla hélix do ADN que é constituído por vários pares complementares de nucleótidos. Cada nucleótido é por sua vez constituído por uma base nucleotídica, um açúcar e um grupo fosfato.

Em virtude das bases nucleotídicas provirem da purina ou da pirimidina, elas classificam-

se em conformidade, i.e. bases púricas e pirimidínicas.

As bases nucleotídicas derivadas da purina são a Adenina e a Guanina, enquanto que as pirimidínicas serão a citosina e a timina (no MA há a pequena diferença da timina ser substituída pelo Uracilo). O segundo elemento que entra na constituição de um nucleótido do ADN é o açúcar que se designa por desoxiribose. A ligação entre a base nucleotídica (já vimos que apenas pode ser ou a adenina, ou a guanina, ou a citosina, ou a timina no ADN) e a desoxiribose ocorre entre o átomo de nitrogénio na posição 9 de uma base púrica ou o localizado na posição 1 de uma base pirimidínica e o átomo de carbono na posição 1 da desoxiribose (quer dizer do açúcar), realizando assim a ligação N-glicosídica.

O terceiro elemento constituinte de um nucleótido é o grupo fosfato que se liga ao terminal hidroxilo do açúcar.

A “montagem” de uma “fita” de ADN realiza-se quando o grupo hidroxilo localizado no açúcar de um nucleótido liga-se ao grupo fosfato também ligado ao grupo hidroxilo do açúcar pertencente ao nucleótido a seguir e assim sucessivamente.

Uma única “fita” de ADN é por definição um polímero nucleotéico. A junção das duas fitas que, ao enrolarem-se em hélix vão formar a molécula completa do ADN, não ocorre por acaso.

Necessariamente deve haver uma complementaridade entre uma base nucleotídica em cada posição de uma “fita” e a outra base correspondente na “fita gémea”. Assim, se numa “fita” estiver a adenina, na fita complementar deve estar a timina e se numa posição estiver a guanina na complementar deve estar a citosina. É contra-natura a ligação de pares entre timina e guanina ou entre a citosina e a adenina (normalmente tem de ser: adenina → timina e guanina → citosina ou vice versa). Desta forma, uma molécula de ADN tem um número duplo de nucleótidos que estiveram presentes em cada uma das duas “fitas” que a compõem (porque os nucleótidos dispõem-se transversalmente aos pares ligados por pontes não covalentes de hidrogénio) [6,7,9].

Nas células eucariotas existe o ADN nuclear e extra-nuclear. Nesta última localização existe apenas nas mitocôndrias e exprime-se de

forma abreviada mtADN. Sempre que o ADN não estiver antecedido do prefixo mt, supõe-se que está a tratar-se do ADN nuclear. Este, existe no seio dos 46 cromossomas que constituem o cariotipo humano (em outras espécies animais e plantas o cariotipo é qualitativa e quantitativamente diferente), sendo que o cromossoma sexual Y é o que menos ADN contém e consequentemente menos genes.

Um grupo de três pares contíguos e complementares de nucleótidos designa-se por **codon** e este codon constitui a base para a síntese de um aminoácido nos ribossomas da respectiva célula. Este processo que consiste na conversão de um tripleto do ADN (três pares de nucleótidos) num aminoácido nos ribossomas do citoplasma da célula, constitui o fundamento do funcionamento da informação genética.

O processo da informação genética é constituído primeiro pela transcrição, que é a conversão de uma fita de ADN numa de ARN mensageiro (mRNA), seguida da tradução do ARN em ARN ribossómico e de transporte (respectivamente rARN e tARN), os quais vão finalmente tornar possível a síntese de um aminoácido no complexo ribossómico citoplasmático da respectiva célula.

Como se pode calcular (teoria matemática de arranjos e combinações), existem 64 possibilidades de se constituírem tripletos a partir das quatro bases nucleotídicas do DNA (adenina, guanina, citosina e timina), porém

existem apenas 20 aminoácidos essenciais, que constituem as cadeias polipeptídicas (quer dizer vários aminoácidos). Por esta razão é que um aminoácido pode ser codificado por mais de um tripleto, porque há uma redundância de tripletos (teoricamente são 44 a mais) para o número de aminoácidos essenciais existentes [4].

O conjunto de 64 tripletos existentes do ADN, é o que se designa por **código genético**. Uma vez que um gene corresponde a um segmento de ADN com um número variável de nucleótidos, quer dizer que esse número é próximo ou exactamente múltiplo de três, porque um gene tem necessariamente vários codons.

Por outro lado, um gene (que é a expressão de vários codons) está na base da transcrição que vai levar a síntese de um polipeptídeo (quer dizer cadeia de vários aminoácidos); uma proteína cuja estrutura requer várias cadeias polipeptídicas, deve ser produzida por vários genes. Assim basta que um desses genes tenha alguma mutação, para a respectiva proteína apresentar igualmente uma alteração. É o caso, por exemplo, da hemoglobina S. cujo codon GAG (guanina+adenina+guanina) localizado no cromossoma 11 do aminoácido da posição 6 do polipeptídeo β está alterado para GTG (guanina+timina+guanina), provocando as alterações clínicas devastadoras que conhecemos nos drepanocíticos.

Código Genético — transcrito no ARN mensageiro (ARNm) [6]

BASES NUCLEOTÍDICAS					
1ª base	2ª base				3ª base
URACILO	URACILO	CITOSINA	ADENINA	GUANINA	Uracilo Citosina Adenina Guanina
	Fenilalanina	Serina	Tirosina	Cisteína	
	Fenilalanina	Serina	Tirosina	Cisteína	
	Leucina	Serina	Stop	Stop	
CITOSINA	Leucina	Serina	Stop	Triptofano	Uracilo Citosina Adenina Guanina
	Leucina	Prolina	Histidina	Arginina	
	Leucina	Prolina	Histidina	Arginina	
	Leucina	Prolina	Glutamina	Arginina	
ADENINA	Leucina	Prolina	Glutamina	Arginina	Uracilo Citosina Adenina Guanina
	Isoleucina	Treonina	Aspargina	Serina	
	Isoleucina	Treonina	Aspargina	Serina	
	Isoleucina	Treonina	Lisina	Arginina	
GUANINA	início(met)	Treonina	Lisina	Arginina	Uracilo Citosina Adenina Guanina
	Valina	Alanina	Ac. Aspartico	Glicina	
	Valina	Alanina	Ac. Aspartico	Glicina	
	Valina	Alanina	Ac. Aspartico	Glicina	

Genoma, genómica, proteoma e proteómica

Em *stritu sensu* o conceito de **Genoma** corresponde ao acervo total de genes presentes nos cromossomas duma célula eucariota e haplóide; contudo em senso lato, o termo tem sido também empregue para designar o acervo total de genes presentes nos cromossomas de organismos procariotas ou em células diplóides (somáticas) de eucariotas.

O estudo integral dos genomas dos seres vivos, e.g. eucariotas, procariotas e plantas, é designado por genómica. Esta é mais abrangente que a genética na medida em que a genética trata fundamentalmente dos mecanismos e conseqüências da hereditariedade e a genómica (conceito introduzido em 1987 por Mckusick e Ruddle) ocupa-se de todos os aspectos da biologia determinados pela actividade dos genes, comparação de genomas de diferentes seres vivos, mapeamento, sequenciamento, evolução, análise funcional dos genes, etc. [9,11,12,14].

O estudo do grupo de famílias de proteínas que compõem os seres vivos é designado por proteómica e o respectivo substrato é o **proteoma**.

O ADN de uma célula germinativa no ser humano (o óvulo ou o espermatozóide), tem em media 3×10^9 pares de bases nucleotídicas nos seus 23 cromossomas (por ser uma célula haplóide).

A célula procariota da *Escherichia coli* tem apenas 4.639.221 pares de bases (pb abreviatura em inglês) nucleotídicas, organizadas em 2657 genes funcionais (62%) e 1632 genes sem função conhecida. As plantas têm geralmente um genoma muito maior, como por exemplo, o trigo, o arroz e o milho, que apresentam genomas que variam de $5-17 \times 10^9$ pb.

Se compararmos um par de bases a uma letra do alfabeto normal da nossa língua e que a mesma meça apenas 1mm, o total de 3.000.000.000 pb do genoma humano (também representa-se 3000 Mb, o que significa três mil megabases) corresponderia a um texto com 3000 Km de comprimento e ocuparia um espaço de 750 MB no disco duro de um micro computador.

Contudo, o proteoma do ser humano e outros mamíferos não requer tanto acervo genético e

apenas 30% do genoma humano é constituído por genes com relevância na sua síntese. Os 70% residuais do ADN humano é que constituem o ADN repetitivo. Em função da extensão e estrutura de pares de bases repetidas, estes classificam-se em diversas formas, dentre as quais assumem significado em Patologia humana, o surgimento de satélites (repetições constituídas por 100-6000 pb), minisatélites (10-20 pb) e microsatélites (2-5 pb).

As hemofilias A e B bem como os carcinomas do cólon e recto são com frequência a expressão de mutações e instabilidade em microsatélites.

Estudo Dos Genes

O estudo dos genes é feito por intermédio do mapeamento, sequenciamento, hibridização, clonagem e análise global do acervo genético. O mapeamento elucida a posição dos genes e a relação destes com outros genes num mesmo cromossoma (uma espécie de anatomia topográfica no cromossoma) [7,8]; o sequenciamento estabelece a sucessão dos genes no ADN; a hibridização é a junção de um segmento de ADN a um outro, cuja composição de genes conhecemos previamente enquanto que a clonagem é a replicação de segmentos ou a totalidade da molécula do ADN, produzindo muitas outras "fitas" de cada uma das duas "fitas" originais, que finalmente voltam a hibridizar-se (juntar-se) entre si, em observância do princípio da complementaridade das bases nucleotídicas. Existe uma relativa diversidade de métodos para realizar qualquer uma das abordagens acima descritas, que levam em conta a biologia molecular, a citogenética e o recurso a microcomputadores de grande capacidade, não nos parecendo justificado entrar em detalhes sobre os procedimentos dos mesmos [12].

O elevado benefício potencial diagnóstico e terapêutico que a genómica pode oferecer à medicina bem como o recurso à mesma, para estudos nos domínios de outras ciências humanas e naturais no geral, determinaram o surgimento do **Projecto do Genoma Humano**, de realização multicêntrica e com implicações que elevarão as ciências biológicas para um paradigma que muito poucos antes previram.

PROJECTO DO GENOMA HUMANO

O projecto do genoma humano tem como objectivos o mapeamento e sequenciamento de genes e marcadores (exons, introns, etc), comparação com o genoma de outros seres vivos e a análise da enorme quantidade de dados que vão estando disponíveis.

O projecto integra vários países e organizações privadas, tendo sido constituídos cinco principais centros de concentração dos dados dos quais o Instituto de Investigação sobre o Genoma Humano nos Estados Unidos da América, desempenha o papel de agência central [14,15]

O conhecimento que for sendo acumulado a medida que o genoma humano for sendo gradualmente mapeado e sequenciado, trará benefícios para a humanidade nos domínios da origem e evolução da humanidade, paleontologia, outras ciências humanas e biológicas [16,17,18], interessando-nos no contexto deste trabalho o domínios da medicina.

Os benefícios em medicina traduzir-se-ão em melhores abordagens (i) diagnósticas, desde que se conheça o gene (ou os genes) que é responsável por uma determinada patologia, (ii) terapêutico e (iii) prognóstico: a avaliação prognóstica de determinadas doenças com base na presença ou ausência de um determinado gene, já é prática corrente em oncologia, hemopatias, etc.

A terapêutica génica consiste na transferência, para o interior de células portadoras de genes mutantes, de genes normais. Do ponto de vista de procedimentos ou técnicas de transferência de genes, são usados os seguintes: (i) utilização de vectores virais, nos quais tenha sido inserido o gene normal que se pretende transferir, uma vez que os vírus apresentam predilecção para determinadas células (tropismo celular), o que permite infectar de forma selectiva as células produtoras de uma determinada substância no organismo. Este é uma técnica em exploração para se resolver o problema da terapêutica génica da drepanocitose, na medida em que os métodos actualmente disponíveis, e.g. o emprego da hidroxiureia ou o transplante da medula óssea, não têm sido muito efectivos; (ii) o emprego de retrovirus é apenas uma forma variante para se procurar atingir sobretudo as células linfóides; (ii) o emprego

de lipossomas e de (iii) substâncias químicas (fosfato de cálcio) como veículos dos genes, pela capacidade que essas estruturas tem de penetrar a membrana celular das células, nas quais existe o gene mutante. O primeiro caso beneficiário da terapêutica génica e que tem entusiasmado a comunidade médica, foi o da senhora Ashanti da Silva, que era portadora de uma imunodeficiência da resposta imunitária celular (células T), que a predispunha a frequentes infecções respiratórias e que os clínicos interpretavam como sendo crises asmáticas ou bronquiolites de causa idiopática. Identificou-se a mutação do gene da adenosina diaminase, que após substituição pelo seu alelo normal, a senhora da Silva está curada da doença.

Contudo, o conhecimento do genoma humano poderá também trazer desafios difíceis de harmonizar, em função dos diferentes grupos de interesse, e.g. políticos, gestores de saúde, gestores de recursos humanos, legisladores, empresários, religiosos, etc.

Muitas doenças provenientes de mutações estruturais (isocromossomas, translocações, deleções, inversões, etc) e quantitativas (poliploidias e aneuploidias) do cariótipo bem como algumas provenientes de mutações génicas com implicações no proteoma humano, já são de domínio corrente. Muitas outras, ainda conhecidas sob a designação de doenças idiopáticas, degenerativas, metabólicas, etc (hipertensão arterial, tesaurismoses, etc), deixarão certamente de o ser, tão breve estejam mapeados os genes cuja estrutura ou função alterada seja o *primum movens* da cascata etiopatogénica explicativa das alterações somáticas e clínicas.

Mesmo em contextos epidemiológicos como o nosso, severamente açotados por doenças infecciosas potenciadas pela pobreza e pelo subdesenvolvimento, parece-nos importante conferir alguma atenção a essas novas fronteiras da ciência.

A patologia de etiologia endógena tem e terá de forma crescente relevância em saúde pública. A drepanocitose, albinismo, cancro, hipertensão arterial, diabetes, etc., contribuem de forma pesada para a diminuição da longevidade, o sofrimento e o aumento da pobreza dos cidadãos e agregados familiares. Por outro lado, ao longo da história da

humanidade, esta nunca esta esteve tão próxima de decifrar, empregar em seu benefício ou manipular os fundamentos mais íntimos da vida do ser humano e de outros seres vivos, desde os vírus às baleias, ou de pequenas plantas a enormes árvores conhecidas em muitos ecossistemas.

Contudo o avanço da ciência normalmente traz consigo benefícios, mas pode também trazer certos malefícios.

Com o mapeamento e sequenciamento do genoma humano, tornam-se crescentes os desafios de índole ética, social e legal [13] que visem impedir eventuais práticas de

eugenia, discriminação e aumento de tensões sociais ou outras manipulações ao serviço dos mais ignominiosos interesses.

Agradecimentos

Agradeço aos Dr. Fortunato Silva e Prof. Dr. Rui Pinto pelo convite que se dignaram formular-me para apresentar este artigo no primeiro número da Revista Científica da Clínica Sagrada Esperança e pela perseverança em aguardar que o tivesse terminado.

REFERÊNCIAS

- 1 - Ricki L.: Human Genetics, Concepts and applications. 6th ed. McGraw-Hill, New York, 2005.
- 2 - Weiling F: Johann Gregor Mendel: Der Mensch und Foster. II Teil. Der Ablauf der Pisum Versuche nach der Darstellung. Med. Genetik 2:208-222. 1993.
- 3 - Griffiths AJF. : An introduction to genetic analysis. 7th ed. W.E. Freeman & Co. New York, 2000
- 4 - Vogel F, Motulsky AG.: Human Genetics. Problems and approaches. 3th ed., Springer Verlag, Heidelberg-New York, 1997
- 5 - Komberg RD, Lorch Y.: Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98: 285-294, 1999.
- 6 - Lodish H. et al. Molecular cell biology. 4th ed. W.H. Freeman & Co., New York, 2000
- 7 - Tyler-Smith C., Willard HF. :Mammalian chromosome structure. Curr. Opin. Genet. Dev.:3:390-397, 1993
- 8 - Miller OJ., Therman E.: Human Chromosomes. 4th ed., Springer Verlag, New York, 2001.
- 10- Albeit B., et al.: Essential Cell Biology. Garland Publishing Co. New York, 1998
- 11 - Brent R.: Genomic biology. Cell 100:169-183, 2000
- 9 - Singer M., Berg P.: Genes & Genomes. Blackwell Scientific, Oxford, 1991
- 12 - Adams MD., et al.: The genome sequence of drosophila melanogaster : Science 287 :285- 295, 2000
- 13 - Collins FS: Shattuck Lecture-Medical and societal consequences of the human genome project New Eng. J. Med. 341: 28-37. 1999
- 14 - Green ED, Cox DR, Myers RM.: The human genome project and its impact on the study of human disease, In Scriver CR et al., eds. The Metabolic and Molecular Bases of inherited disease. 7th ed. McGraw-Hill, New York, 1995
- 15 - Brown TA. Genomes. Bios Scientific Publishers, Oxford, 1999. Goodman L.: The human genome project aims for 2003. Genome Res. 8: 997-999, 1998.
- 16 - Rubin GM et al. Comparative genomics of the eukaryotes. Science 287:2204-2215, 2000
- 17 - Yunis JJ., Prakash O.: The origin of man: A chromosomal pictorial legacy. Science 215: 1525-1530, 1982.
- 18 - Caim RL., Stoneking M., Wilson AC. Mitochondrial DNA and human evolution. Nature 325: 31-36, 1987.