



SNPs:

Sutis diferenças de um código

O genoma sob um novo ponto de vista

A medida que a seqüência nucleotídica do genoma humano foi sendo desvendada, uma constatação evidente foi o grande número de variações de ponto encontradas ao se comparar segmentos correspondentes do genoma (figura 1). Antes mesmo do primeiro rascunho da seqüência completa ser divulgado, uma grande parcela da comunidade científica já voltava sua atenção para essas pequenas e abundantes variações dispersas por todo o código genético (figura 2). As mais comuns (ocorrem, aproximadamente, a cada 600 bases) são denominadas polimorfismos de nucleotídeos únicos ou SNPs (pronuncia-se 'S' 'N' 'Ps' ou SNiPs), e correspondem a posições onde existe uma alternância dos nucleotídeos A, C, G e T, em uma freqüência alélica mínima de 1% em uma dada população (Brookes, 1999).

Os SNPs ocorrem tanto em regiões codificadoras como em não codificadoras dos genomas. Em regiões codificadoras, quando resultam em uma substituição de aminoácido na seqüência protéica, são denominados não sinônimos, podendo a substituição ser conservativa ou não conservativa em função das características dos aminoácidos envolvidos na troca. Nesses casos, podem haver modificações estruturais e funcionais na proteína. Embora SNPs sinônimos não alterem a seqüência protéica, eles podem modificar a estrutura e a estabilidade do RNA mensageiro,

e, conseqüentemente afetar a quantidade de proteína produzida. Além disso, SNPs podem promover: *splicing* alternativo, alterações no padrão de expressão de genes (como no caso de alterações em seqüências de promotores), geração ou supressão de códons de terminação ou poliadenilação na molécula de RNA mensageiro e alteração nos códons de iniciação de tradução.

As substituições mais freqüentes que ocorrem no DNA são as que

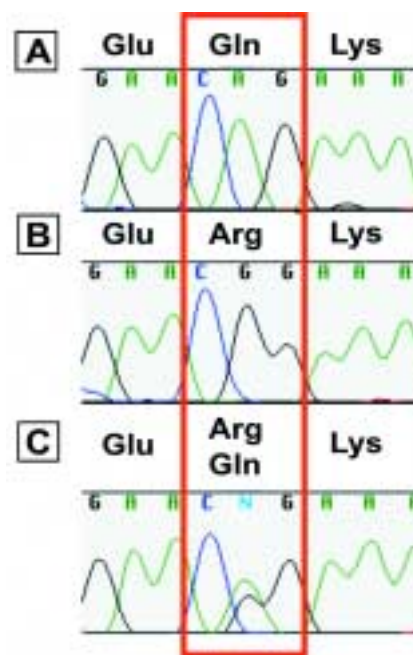


Figura 1 - Alinhamento de seqüências correspondentes ao mesmo segmento genômico de três indivíduos, mostrando um SNP não sinônimo A/G, que causa a substituição de glutamina para arginina na seqüência protéica.

- A) Indivíduo homocigoto AA.
- B) Indivíduo homocigoto GG.
- C) Indivíduo heterocigoto AG

Pedro Edson Moreira Guimarães
Orientação do Prof. Dr. Emmanuel Dias Neto
Laboratório de Neurociências – LLM27
Departamento e Instituto de Psiquiatria
Universidade de São Paulo
pedson@usp.br

Maria Cristina Ramos Costa, PhD
Coordenação do Prof. Dr. Marco Antonio Zago
Centro de Terapia Celular – Hemocentro de Ribeirão Preto
Faculdade de Medicina de Ribeirão Preto
Universidade de São Paulo
mrcosta@usp.br

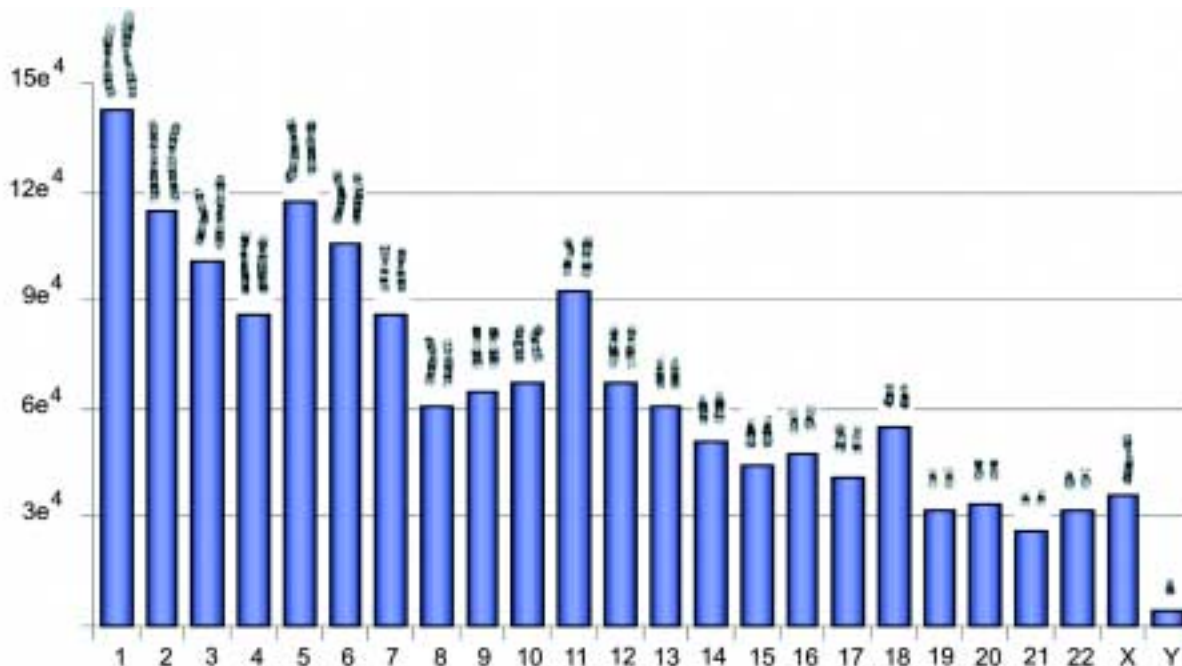


Figura 2 - Distribuição por cromossomo dos SNPs depositados no banco de dados dbSNP (<http://www.ncbi.nlm.nih.gov/SNP>). Aqui podemos observar, como mencionado no texto, que os polimorfismos humanos estão dispersos por todo o genoma

envolvem bases nitrogenadas de mesma característica estrutural, ou seja, trocas entre duas purinas (A/G ou G/A) ou duas pirimidinas (C/T ou T/C) e são denominadas transições. As transversões são substituições de uma purina por uma pirimidina ou vice-versa. Essas alterações, algumas vezes, têm origem em erros de incorporação de bases durante a replicação do DNA; em outros casos, são causadas por lesões no DNA por agentes ambientais. Caso essas mutações ocorram em células germinativas, sejam transmitidas às gerações seguintes e se fixem na população em uma frequência mínima de 1%, passam a ser denominadas de polimorfismos (Kwok & Gu, 1999).

Atualmente, os SNPs podem ser empregados nas mais diversas áreas, como medicina forense, antropologia molecular, evolução, definição de marcadores de predisposição a determinadas patologias e de prognóstico a diferentes tratamentos, genética de populações, conservação e manejo de fauna, farmacogenética, desenvolvimento de vacinas, entre outras. Como exemplo, podemos citar o mecanis-

mo de resistência a uma droga recentemente desenvolvida para o tratamento da leucemia mieloide crônica, o STI571, que é um inibidor específico da atividade tirosina quinase da proteína híbrida BCR-ABL, responsável pelo desenvolvimento dessa neoplasia. Apesar dos resultados favoráveis obtidos com essa nova droga, um grupo de pacientes mostrou-se resistente a ela. Em um estudo recente, foi demonstrado que a resistência à droga está relacionada com um polimorfismo no resíduo 315 da proteína. A substituição de treonina por isoleucina nessa posição impede a formação de uma ponte de hidrogênio essencial para a ligação do inibidor. Dessa forma, uma análise prévia desse polimorfismo em pacientes pode indicar quais devem ser tratados com o STI571 (Gorre *et al.*, 2001).

Um outro estudo recente, conduzido por um grupo de pesquisadores brasileiros, identificou um SNP não sinônimo no gene COL18A1, que eleva em duas vezes e meia o risco de se desenvolver câncer de próstata (Iughetti *et al.*, 2001). A substituição Asp104Asn na endostatina, um produto do

colágeno XVIII e potente inibidor da angiogênese, prejudica a função dessa proteína, o que explica a maior predisposição ao desenvolvimento do tumor.

Em busca de outros polimorfismos que pudessem contribuir para o desenvolvimento de abordagens farmacológicas individualizadas, ou seja, que atendessem ao perfil genético de cada paciente, foi criado, em 1999, um consórcio envolvendo onze gigantes da indústria farmacêutica e de biotecnologia, denominado *The SNP Consortium*. O projeto se estendeu até o final de 2001 e utilizou um orçamento estimado em 50 milhões de dólares, tendo sido capaz de identificar, juntamente com o Projeto Genoma Humano, 1,42 milhão de SNPs no genoma humano (Sachidanandam *et al.*, 2001).

Uma iniciativa anterior, lançada em 1996 pelo *National Cancer Institute* dos USA e denominada *Cancer Genome Anatomy Project* (CGAP), visa gerar informações e tecnologias necessárias para decifrar a anatomia molecular das células tumorais. Dentre as ferramentas desenvolvidas, algumas são volta-

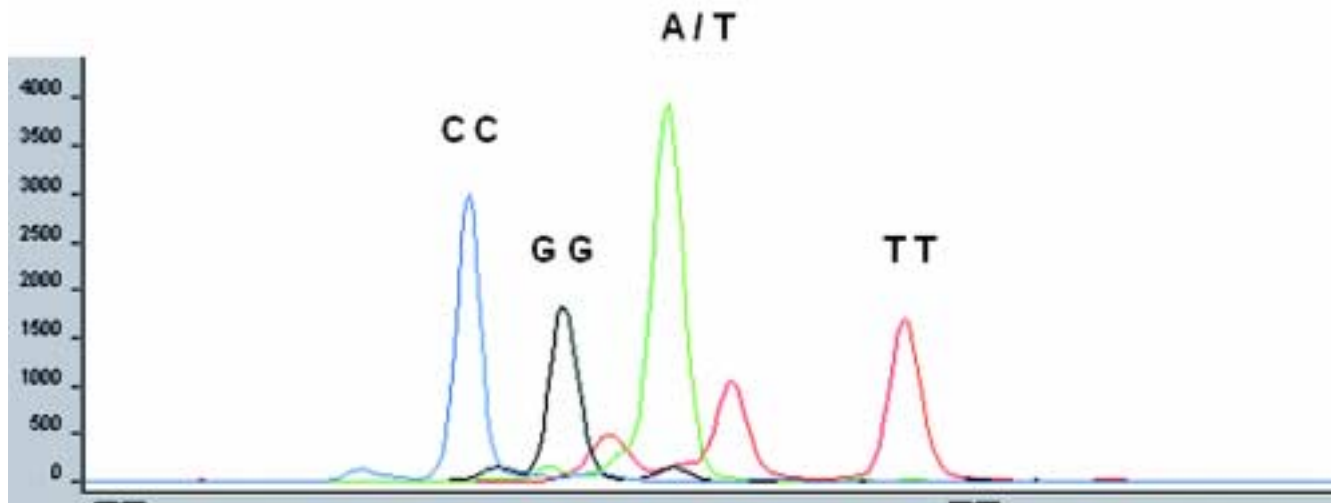


Figura 3 - Genotipagem em sistema de multiplex. Aqui um único indivíduo foi genotipado para quatro diferentes SNPs. Podemos ver que o indivíduo genotipado é homocigoto CC para o SNP1, homocigoto GG para o SNP2, heterocigoto AT para o SNP3 e homocigoto TT para o SNP4

das para a identificação e mapeamento de SNPs. Como resultado desse projeto, foram identificados, até o momento, mais de 20 mil SNPs humanos. As informações de mapeamento e caracterização de cada polimorfismo estão disponibilizadas em bancos de dados públicos (<http://gai.nci.nih.gov>, <http://www.ncbi.nlm.nih.gov/SNP>).

No Brasil, utilizando as seqüências geradas pelo Projeto Genoma

do Câncer Humano (FAPESP/LICR-HCGP), encontra-se em andamento uma iniciativa de identificação de SNPs na região codificadora de genes humanos (HCGP-SNP, http://bit.fmrp.usp.br/hcgp_snp). Os dados produzidos até o momento resultaram na identificação e validação experimental de novos SNPs humanos, que estão sendo avaliados em população normal para determinação de frequência alélica e em amostras derivadas de diferentes tumores.

É importante ressaltar que nos projetos de identificação de SNPs, principalmente aqueles em larga escala, a participação de uma equipe de bioinformática é fundamental. Graças ao desenvolvimento tecnológico de *softwares* e de equipamentos cada vez mais robustos e sofisticados, vem sendo possível gerar e analisar um grande número de dados de seqüências de DNA. Inicialmente, as abordagens de detecção de SNPs consistiam em amplificar fragmentos genômicos equivalentes, de regiões gênicas candidatas a determinado fenótipo (predisposição a uma doença neuropsiquiátrica, por exemplo), do DNA de vários indivíduos, entre grupos teste e controle, e comparar suas seqüências buscando variações. Atualmente, a maior parte dos estudos desenvolvidos visan-

do a identificação de SNPs em larga escala tem explorado as milhares de seqüências presentes nos bancos de dados, incluindo clones genômicos e, principalmente, seqüências de cDNA ou ESTs (*Expressed Sequence Tags*). Os bancos de dados de ESTs apresentam um elevado grau de redundância, e o fato de as seqüências serem derivadas de tecidos de vários indivíduos torna a utilização desses bancos muito promissora para a identificação de polimorfismos.

As ferramentas de bioinformática são utilizadas em várias etapas do estudo de polimorfismos, desde a identificação das variações até a predição do efeito delas. Outra aplicação importante é a construção e manutenção de bancos de dados e de ferramentas que podem ser acessadas pela internet. Esses bancos atuam como sedes de referência para a deposição de SNPs, sendo que no maior banco público de polimorfismos, que é mantido pelo *National Center for Biotechnology Information* (dbSNP, <http://www.ncbi.nlm.nih.gov/SNP>), já estão depositados mais de 4,1 milhões de SNPs humanos e um número considerável de polimorfismos de outras espécies.

Apesar da indiscutível contribuição da bioinformática na identificação de SNPs, a necessidade de

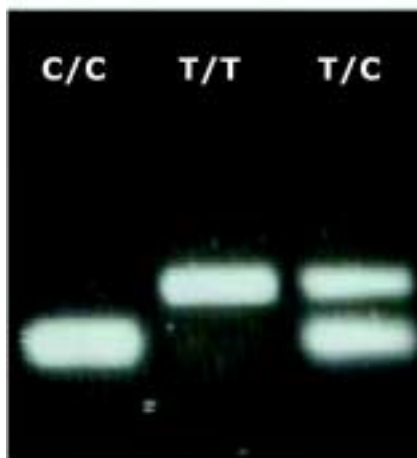


Figura 4 - Exemplo de análise de um polimorfismo por RFLP, onde a presença do nucleotídeo C cria um sítio de restrição para a enzima de restrição. Gel de agarose contendo na raia 1 um indivíduo homocigoto CC, na raia 2 um homocigoto TT e na raia 3 um heterocigoto TC (apenas o alelo C é digerido)

validação experimental dos dados e estudos de associação são indispensáveis para se avaliar o potencial informativo de cada polimorfismo. Dessa forma, o surgimento e a evolução de métodos de validação e de genotipagem de amostras para SNPs confirmados também merecem reconhecimento. Hoje, é possível genotipar vários SNPs em várias amostras em um mesmo tubo de reação; como exemplo podemos citar o sistema de multiplex desenvolvido pela *Applied Biosystems™*, que pode genotipar até 10 SNPs em uma mesma reação (figura 3). Outros métodos robustos de genotipagem podem ser citados: *chips* de DNA ou *microarrays*, cromatografia líquida de alta pressão desnaturante (DHPLC), espectrometria de massa, PCR em tempo real, mini-sequenciamento e sequenciamento de uma única base. Entretanto, em análises de menor escala, métodos como sequenciamento (figura 1) e digestão por enzimas de restrição – RFLP (*Restriction Fragment Length Polymorphism*) são rotineiramente empregados. Na figura 4, é mostrado um exemplo de SNP genotipado utilizando uma enzima de restrição, onde a presença do polimorfismo cria um sítio para a enzima, podendo os diferentes alelos ser identificados em um simples gel.

Perspectivas

Um tema ainda bastante polêmico que envolve polimorfismos de DNA é a legalidade de se patentear dados de SNPs. De um lado, há o interesse econômico de grandes empresas farmacêuticas. Um dos objetivos é o de conhecer o genoma humano, de modo a associar SNPs com resposta a drogas. De outro lado, os pesquisadores e as instituições envolvidos na geração de dados públicos. Muitas vezes, as empresas farmacêuticas utilizam os dados públicos em seus projetos, e deles retiram os resultados que serão transformados em patentes privadas sem retorno aos pes-

quisadores e às instituições responsáveis pela produção desses dados. Essa discussão ainda está longe de alcançar um ponto final, mas uma tendência que vem-se tornando consenso entre os envolvidos é a possibilidade de se patentear produtos industrializados específicos criados a partir de dados de SNPs, mas não os polimorfismos em si.

Até o momento, grande parte dos estudos tem relatado a associação entre um único SNP e uma determinada patologia, ou mesmo definido um SNP como um marcador para predisposição a uma doença. Entretanto, visando a compreensão das bases moleculares de doenças de características genéticas mais complexas, pesquisas têm sido intensificadas no sentido de tentar associar, ao fenótipo alterado, conjuntos de determinados SNPs, denominados haplótipos (Brookes, 2002). Embora esse tipo de enfoque necessite de maiores investimentos tanto tecnológicos como financeiros e envolva um número maior de amostras, ele constitui uma abordagem muito promissora nas patologias humanas mais comuns, como distúrbios neuropsiquiátricos, doenças autoimunes, doenças cardiovasculares e câncer.

Podemos vislumbrar para um futuro não muito distante a compreensão e subseqüentes tratamentos para essas patologias. Teremos uma medicina pessoal, voltada para as características genéticas de cada indivíduo. A indústria farmacêutica já busca, através de drogas mais específicas, suprir essas necessidades individuais, conseguindo, com algum sucesso, reduzir os efeitos colaterais de vários medicamentos. Testes mais acurados estão sendo desenvolvidos e será possível definir, antes mesmo do nascimento, um mapa clínico do genoma de cada pessoa, contendo informações como as patologias a que será mais susceptível e a quais tratamentos responderá melhor. Sob essa nova ótica, não é arriscado dizer que as

pessoas poderão viver mais e com melhor qualidade de vida.

Agradecimentos

Somos especialmente gratos ao Prof. Dr. Emmanuel Dias Neto pelo apoio e pela revisão do manuscrito. Agradecemos à FAPESP as bolsas de pós-doutorado de Maria Cristina Ramos Costa e a bolsa de mestrado de Pedro Edson Moreira Guimarães.

Referências

- Brookes AJ. 1999. The essence of SNPs. *Gene* **234**(2):177-186.
- Brookes AJ. 2002. 4th International meeting on Single Nucleotide Polymorphism and complex genome analysis various uses for DNA variations. *Eur J Hum Genet* **10**(2):153-155
- Gorre ME, Mohammed M, Ellwood K, Hsu N, Paquette R, Rao PN, Sawyers CL. 2001. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* **293**(5531):876-880.
- Iughetti P, Suzuki O, Godoi PH, Alves VA, Sertie AL, Zorick T, Soares F, Camargo A, Moreira ES, di Loreto C, Moreira-Filho CA, Simpson A, Oliva G, Passos-Bueno MR. 2001. A polymorphism in endostatin, an angiogenesis inhibitor, predisposes for the development of prostatic adenocarcinoma. *Cancer Res.* **61**(20):7375-7378.
- Kwok PY, Gu Z. 1999. Single nucleotide polymorphism libraries: why and how are we building them? *Mol Med Today* **5**(12):538-543.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, *et al.* 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409** (6822): 928-933. 