

Descobrendo Genes no Século XXI: Enfoque na Área de Onco-Endocrinologia

revisão

RESUMO

Patricia Dahia

A área de endocrinologia genética e oncológica tem sido alvo de enorme avanço nos anos recentes. A descoberta de genes responsáveis por neoplasias hereditárias nas últimas décadas representou uma fonte importante de informações concernentes à avaliação de risco, prevenção e aconselhamento genético. Neste particular, a caracterização do gene responsável pela neoplasia endócrina múltipla tipo 2 (MEN 2) representou a "pedra angular" para o desenvolvimento do campo de oncogenética clínica. Portanto, estima-se que o estudo genético e funcional das neoplasias hereditárias será igualmente capaz de liderar o avanço para um melhor manuseio clínico e terapêutico destas doenças. Um dos grandes desafios do futuro reside em se entender o intrincado mecanismo de interação entre genes e proteínas no controle do desenvolvimento e regulação dos diferentes sistemas biológicos. Uma das conseqüências mais antecipadas deste avanço é a sua aplicação para o tratamento específico e "individualizado" de todas as doenças humanas. Esta breve revisão tratará de aspectos gerais que envolvem o processo de identificação de novos genes e sua associação com condições clínicas específicas. As técnicas clássicas de clonagem serão apresentadas ao lado de estratégias modernas de identificação e análise de genes. O papel da bioinformática no Projeto Genoma Humano e o imenso potencial que esta informação traz para acelerar o processo de caracterização de novos grupos de genes serão brevemente discutidos. (**Arq Bras Endocrinol Metab 2002;46/4:341-350**)

Descritores: Oncogenética clínica; Neoplasias hereditárias; Neoplasia endócrina múltipla tipo 2; Bioinformática; Projeto Genoma Humano

*Department of Cancer Biology,
Dana-Farber Cancer Institute,
Harvard Medical School, Boston,
MA, EUA.*

ABSTRACT

Discovering Genes in the XXI Century: Focus in the Onco-Endocrinology Field.

The field of endocrine and cancer genetics has grown remarkably in the past few years. The discovery of genes responsible for inherited forms of cancer has already given us powerful tools for cancer risk assessment, pre-morbid prediction and genetic counseling. The discovery of the susceptibility gene for multiple endocrine neoplasia type 2 (MEN 2) was fundamental in developing the field of clinical cancer genetics. It is hoped that further genetic and functional studies of hereditary cancers, including inherited endocrine neoplasia syndromes, will also lead the way to improved clinical management ability in prevention and treatment areas alike. One of the greatest challenges that lie ahead is to understand the intricate mechanism of gene-protein interactions that control development and regulation of multiple biological systems. One of the most anticipated consequences of this progress is that it will bring further insight and tools that will also enable full translation into therapeutic options that can be tailored to patients according to their most specific "molecular requirements". This brief review covers general aspects of the gene discovery process and further association with specific disease phenotypes. Traditional cloning techniques will be listed side by side with

*Recebido em 25/05/2002
Aceito em 31/05/2002*

modern strategies for gene discovery and analysis. The role of bioinformatics in the progress of the Human Genome Project and the enormous potential that derives from this information towards the identification of novel genes will be addressed. (Arq Bras Endocrinol Metab 2002;46/4: 341-350)

Keywords: Clinical cancer genetics; Inherited neoplasias; Multiple endocrine neoplasia type 2; Bioinformatics; Human Genome Project

TÉCNICAS DE CLONAGEM GÊNICA: UMA VISÃO GLOBAL

A COMPLETA COMPREENSÃO da base molecular de doenças humanas, que compreende a identificação dos genes primariamente envolvidos, além de sua caracterização funcional, é um dos grandes objetivos da medicina molecular. O progresso acentuado dos últimos anos, largamente relacionado à identificação de genes, já transformou de maneira permanente a nossa percepção da fisiologia e a patologia de diversas condições clínicas. Este conhecimento tem se traduzido até o momento essencialmente em um melhor arsenal diagnóstico, uma capacidade mais acurada de mensuração de riscos associados a doença e também de seguimento pós-tratamento. Certamente, os próximos anos trarão uma aplicação altamente antecipada: a incorporação destas noções para o desenvolvimento de novas drogas, seguida de seu emprego na prática médica.

Alguns dos métodos utilizados para a clonagem e identificação de novos genes possuem um espectro amplo de aplicações, embora as estratégias usadas para este fim possam ser, ocasionalmente, específicas a depender de particularidades da patologia em estudo. Algumas técnicas, por exemplo, são empregadas com extremo sucesso em patologias de natureza hereditária, enquanto outras se restringem ao uso em condições

esporádicas. Como se tornará mais claro nas subseqüentes partes desta revisão, o termo “clonagem” muitas vezes será substituído pela expressão “identificação” de novos genes. Este último termo reflete de forma mais acurada o processo de caracterização de uma nova “estrutura codificadora de proteína” já que, cada vez mais freqüentemente, tal processo prescinde do evento experimental envolvendo clonagem na sua forma literal. Os bancos de dados correntes possuem um grande número de genes tecnicamente listados como unidades gênicas e, portanto, considerados para alguns efeitos “clonados”, porém sem qualquer caracterização de associação funcional. Para efeito desta revisão, estes genes ainda estão por ser “identificados”.

Clonagem posicional

O conceito básico de clonagem posicional compreende a identificação inicial da localização cromossômica, onde se suspeita residir o gene associado a um dado fenótipo (1,2). Os passos subseqüentes envolvem a redução da área crítica ao menor tamanho possível e por fim a identificação do gene-alvo e mutações causadoras da doença.

Os genes responsáveis por um significativo grupo de neoplasias endócrinas hereditárias ou esporádicas foram estabelecidos através do uso de clonagem posicional. Entre estes se encontram o gene *RET* responsável por MEN 2A, 2B e carcinoma medular de tireóide familiar (3); o gene *MEN1* (menin) na síndrome de MEN 1 (4); o gene *VHL* na síndrome de von Hippel-Lindau (5), o gene *PTEN* na síndrome de Cowden e hamartomas relacionadas (6), o gene *PRKARIA* em um dos subtipos de síndrome de Carney (7), entre outros. Uma vez que a “área candidata” a conter o gene é identificada, as técnicas que podem ser usadas para clonar o gene de doenças familiares e esporádicas são virtualmente as mesmas (tabela 1). As estratégias diferem essencialmente na forma como a

Tabela 1. Estratégias utilizadas para identificação dos genes responsáveis por algumas patologias hereditárias com componente endócrino após o mapeamento do gene candidato por análise de ligação.

Gene	Doenças	Técnica de clonagem
<i>RET</i>	MEN 2	Clonagem Posicional, Teste de genes candidatos
<i>MEN1</i>	MEN 1	Clonagem Posicional, Rastreamento de bibliotecas de cDNA, Teste de genes candidatos
<i>PTEN</i>	Cowden syndrome/ Bannayan-Riley-Ruvalcaba syndrome	Clonagem Posicional, Rastreamento de bibliotecas de cDNA, Teste de genes candidatos
<i>VHL</i>	von Hippel-Lindau disease	Clonagem Posicional, Rastreamento de bibliotecas de cDNA, Teste de genes candidatos
<i>PRKARIA</i>	Carney Complex syndrome	Clonagem Posicional, Teste de genes candidatos

área candidata é identificada. Em doenças familiares, a análise de ligação genética (*linkage analysis*) representa a maneira clássica para tal caracterização (8-10). Esta estratégia é discutida abaixo. Este método, no entanto, não pode ser utilizado para a identificação de genes responsáveis por patologias esporádicas. Nestas, em particular em neoplasias esporádicas, a utilização de áreas consistentes de deleção ou anormalidades cromossômicas mais grosseiras em geral são o passo inicial em direção à identificação do gene-alvo (1). Um aspecto interessante a se registrar é a possibilidade, uma vez que uma região candidata é identificada pela análise de ligação de famílias, da aplicação destes dados como base para a caracterização de genes responsáveis por doenças esporádicas. Por exemplo, a caracterização do receptor tirosina quinase RET como o gene causador de MEN 2A levou a sua identificação como responsável pelo desenvolvimento de cerca de um terço de carcinomas medulares de tiróide esporádicos (11).

Análise de ligação genética

O estudo de famílias com susceptibilidade para o desenvolvimento de neoplasias representa um recurso poderoso para o progresso do conhecimento oncológico. Na atual era da medicina molecular, a identificação de uma mutação constitutiva traz consigo dois importantes aspectos de imediato impacto: a) esta informação se traduz na habilidade de se oferecer triagem genética para membros não afetados e sob risco na família; b) torna-se possível estimar o risco de desenvolvimento de tumores. Faz-se necessário, por conseguinte, a implementação de um serviço bem estruturado de aconselhamento genético para apoio a tal sistema de rastreamento.

O estudo de ligação genética é uma estratégia criticamente importante para a identificação de genes responsáveis por patologias hereditárias, que tem sido empregada há décadas com sucesso na caracterização de uma variedade imensa de patologias. Seu uso foi introduzido quando se verificou que variações do DNA humano poderiam ser empregadas como marcadores genéticos em estudos de associações (12). Como resultado dos processos de recombinação durante a divisão celular meiótica, marcadores localizados próximos ao gene da doença revelam uma forte associação com o padrão de distribuição da doença em famílias. O pressuposto básico da análise é que quanto menor a distância entre 2 loci no mesmo cromossomo, maior a chance de que estes alelos estejam “ligados” geneticamente; em outras palavras, que estes alelos são herdados em conjunto em mais de 50% das vezes (50% representaria a probabilidade ao acaso de associação de

dois loci) (13). Uma análise estatística é então aplicada a estes achados para interpretar os resultados (14,15). O objetivo desta análise é estabelecer se dois loci (o locus contendo um marcador e um segundo contendo o gene responsável pela doença) segregam independentemente ou não durante a meiose. Quando os dados sugerem a presença de ligação, verifica-se então que certos alelos segregam exclusivamente com indivíduos afetados. A caracterização de eventos recombinantes específicos dentro da área considerada “ligada” ou associada ao fenótipo em questão permitem a identificação da mínima área cromossômica que abriga o gene responsável pela doença. Esta distância, na prática, pode ser reduzida a alguns milhares de kilobases. A maior aplicação da técnica de ligação é em doenças mendelianas. Tipicamente, de acordo com os critérios de Morton (12), a determinação estridente e inquestionável de ligação requer que seja atingido um *lod score* (logaritmo da probabilidade de associação) superior a 3, ou seja, uma chance menor que 0,001 de que a associação encontrada seja de natureza randômica. No outro extremo, a exclusão de ligação é dada pelo achado de *lod score* inferior a -2. Este último aspecto é altamente relevante, uma vez que no estudo de doenças raras nem sempre é possível atingir um número suficiente de casos para alcançar o limiar de *lod score* desejado. Nestas circunstâncias, a exclusão de certas áreas do genoma em combinação com áreas “sugestivas” de ligação permite a criação de um mapa de ligação acurado.

Embora a identificação da área de associação, também conhecida como mapeamento, apenas ofereça uma referência para a localização do gene que causa susceptibilidade à doença, a sua aplicação pode ser imediata. Mesmo antes que o gene candidato seja identificado, a informação gerada pelo mapeamento já pode ser utilizada para o desenvolvimento de testes de rastreamento baseados na segregação de certos alelos dentro de um grupo familiar onde a doença em questão é transmitida hereditariamente. Por muitos anos, esta estratégia foi utilizada em pacientes com MEN 1 (16), até que o gene causador da doença foi identificado em 1997 (4).

Como mencionado acima, os genes responsáveis por diversas síndromes hereditárias endócrinas foram identificados após o seu mapeamento com estudos de ligação (tabela 1). Ao contrário do cenário favorável com múltiplos exemplos de sucesso do emprego desta estratégia para o mapeamento de novos genes, estudos de ligação são limitados na sua capacidade de direcionar a identificação de genes causadores de doenças multigênicas. Estas doenças

comuns, como hipertensão arterial essencial, diabetes mellitus, artrite, osteoporose, continuam representando um desafio para estudos baseados em análise de grupos familiares (17-19).

Defeitos cromossômicos como base da caracterização de doenças esporádicas

Um importante objetivo na descoberta de novos genes é a identificação de uma região cromossômica associada com um fenótipo específico. Tais anormalidades são em geral pequenas, e com frequência representam apenas eventos secundários e não defeitos primários responsáveis, no caso de neoplasias, pelo crescimento ou progressão tumoral. A análise de perda de heterozigose (*loss of heterozygosity or LOH*, descrita abaixo) representa uma das estratégias mais largamente empregadas para a identificação de áreas candidatas (20,21). Certas anormalidades cromossômicas, ao contrário, podem ser mais aparentes, envolvendo de-feitos estruturais, como translocações, inversões, deleções, ampliações, ou numéricas (perdas ou ganhos de cromossomos inteiros ou regiões cromossômicas). Estas anormalidades são em geral detectadas com o uso de técnicas clássicas de citogenética ou metodologia derivada, como FISH (hibridização *in situ* por fluorescência), SKY (cariotipagem espectral), CGH (hibridização genômica comparativa). Estas técnicas são coletivamente conhecidas como citogenética molecular (22-24). Posteriormente, iremos discutir a versão moderna de uma destas técnicas, que recentemente introduzimos na análise de feocromocitomas, denominada *array CGH*.

Perda de heterozigose (LOH)

Por muitas décadas é conhecido o fato de que dois alelos anormais são necessários para que uma doença recessiva se manifeste. As primeiras observações foram feitas em estudos de pacientes com retinoblastoma por Knudson (25). De acordo com este modelo, duas mutações seriam necessárias para transformar uma célula normal em neoplásica. Em tumores familiares, uma das mutações seria herdada e, portanto, estaria presente nas células germinativas, enquanto que a segunda mutação seria adquirida apenas pelo tecido afetado (mutação somática). No caso de tumores esporádicos, as duas mutações aconteceriam no tecido somático como eventos independentes. Este modelo se tornou conhecido como teoria de Knudson ou teoria dos dois eventos e, desde a sua descrição inicial, foi amplamente confirmado em modelos de neoplasias causadas por genes supressores de tumor, o conceito derivado desta teoria (26-28). Como o primeiro evento é herdado por todas as células no modelo familiar,

isto permite que portadores da mutação sejam rastreados com qualquer tipo de célula: na prática, DNA obtido de leucócitos periféricos é a forma mais comum para representação da célula constitutiva. Uma extensão desta estratégia permite a caracterização do papel de genes supressores também em tumores esporádicos (29-33). Como brevemente mencionado acima, polimorfismos presentes no genoma podem ser utilizados como marcadores de heterozigose e identificação de dois alelos em DNA normal. O estudo comparativo destes marcadores em amostras tumorais permite identificar áreas de deleção pela perda de um dos alelos, ou perda do estado de heterozigose, LOH (34-36).

Com a identificação de milhares de marcadores polimórficos ao longo do genoma, o mapeamento de áreas de deleção tem se tornado progressivamente mais acurado, o que tem facilitado grandemente a identificação de novos genes supressores de tumor. Uma limitação da análise de LOH é a qualidade da amostra tumoral. Células tumorais geralmente são permeadas por células normais adjacentes ou circundadas por vasos que irrigam o tecido neoplásico. O nível de contaminação com estas células normais pode ocasionar resultados imprecisos e equívocos na definição de áreas deletadas, já que estas podem ser “mascaradas” pela presença de células normais. Uma das maneiras de contrapor esta limitação é a microdissecção a laser de regiões histológicas do tumor. Assim, células identificadas como tumorais podem ser identificadas com uma precisão de um mínimo de 6 células dependendo da acurácia do equipamento de microdissecção, e removidas para isolamento de DNA (37).

Hibridação Genômica Comparativa (CGH)

Variações do número de cópias de genes podem ocorrer em um espectro variado de doenças, em especial neoplasias, e são indicativas de alterações em oncogenes e genes supressores de tumor. Assim, a detecção e mapeamento de alterações do número de cópias de certas regiões do genoma oferecem uma importante contribuição para a associação entre defeitos moleculares e o fenótipo estudado.

CGH tem sido largamente empregada no rastreamento ao longo do genoma de defeitos cromossômicos que podem representar a evidência inicial para a identificação de genes associados a doenças neoplásicas (38). Neste método, cromossomos em metáfase são competitivamente hibridados com duas amostras de DNA genômico marcadas diferencialmente: uma amostra a ser testada e um controle normal de referência. Em geral, utilizam-se marcadores

fluorescentes, como Cy3 e Cy5, para estas reações (38). A análise revela a localização cromossômica de ganhos ou perdas medidas pelo tipo de fluorescência detectado. Assim, áreas de ganho de material cromossômico na amostra teste são reveladas por um predomínio do marcador fluorescente desta amostra. Regiões de perda, ao contrário, denotam um predomínio da fluorescência da amostra normal. Áreas inalteradas apresentam uma sobreposição da fluorescência dos dois marcadores (38). Esta técnica tem sido capaz de revelar alterações cromossômicas não identificadas por análise citogenética tradicional (39).

Contudo, o limite de resolução de CGH ainda é baixo, da ordem de 20Mb, em geral, o que restringe a caracterização de defeitos mais sutis. Para resolver esta limitação, foi desenvolvida uma variante técnica de CGH, na qual as amostras teste e controle são hibridadas a uma matriz contendo segmentos de DNA correspondente as diferentes localizações cromossômicas, ao invés das metáfases usadas no método original (39). Esta variante, também conhecida como *array CGH*, é composta de vetores de alta capacidade denominados BACs, que contêm segmentos de cerca de 200kb (ou 0,2Mb) de DNA. Cerca de 30.000 BACs mapeados e representativos do genoma são imobilizados em placa ou lâmina e são usados como a fase sólida para hibridização. O uso destes vetores permite um aumento instantâneo do limite de resolução da técnica de cerca de 10 vezes. Para a detecção de ainda menores alterações cromossômicas, estamos no momento refinando versões específicas de *array CGH* para atingir uma maior capacidade de resolução (40). A figura 1 ilustra a estratégia envolvida nesta técnica.

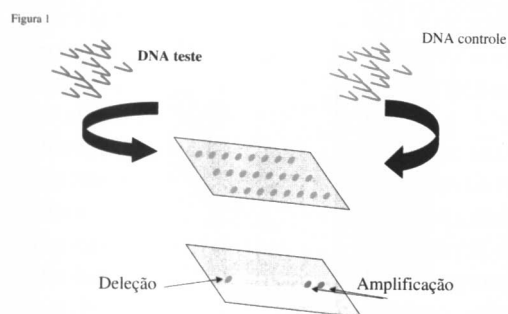


Figura 1. Uso de *array CGH* para identificação de anormalidades no número de cópias de genes ou regiões genômicas. A amostra teste e o controle são marcados diferencialmente por fluorescência e posteriormente hibridadas a uma matriz contendo segmentos de DNA imobilizados. Predomínio da intensidade de fluorescência correspondente à amostra teste ou ao controle são indicativos de ganho ou perda de cópias, respectivamente, na região correspondente da amostra teste.

Estudos baseados em transcrição gênica

Atualmente existe uma gama de técnicas disponíveis para a identificação de defeitos genéticos, que se baseiam na análise da expressão de genes. Algumas destas técnicas se baseiam na comparação das diferenças na abundância de transcritos de RNA entre duas populações celulares como, por exemplo, tumor e tecido normal. Novos genes podem ser identificados através do uso de tais técnicas comparativas. Para este propósito, uma variedade de métodos estão disponíveis: distribuição diferencial (*differential display*) (41), hibridização subtrativa (42), análise seriada de expressão gênica, SAGE (43) e, mais recentemente, estratégias baseadas em chips (*microarrays*) (44). A discussão destas técnicas mais tradicionais não faz parte do escopo deste artigo; muitas revisões tratando do seu uso estão disponíveis na literatura (45,46). Discutiremos a seguir o uso de chips para estudos de expressão gênica em larga-escala, com especial ênfase em uma das muitas aplicações do uso de chips: o seu emprego para a descoberta de novos genes.

Microarrays

Microarrays de cDNA, também conhecidos como DNA *arrays*, DNA chips, biochips e GeneChips, avaliam a expressão de milhares de genes simultaneamente (47). Chips são capazes de fornecer uma ampla fonte de informações, desde a caracterização do perfil de expressão de um determinado tecido durante o desenvolvimento embrionológico até o efeito de drogas na transcrição, genes associados com específicas condições patológicas, subgrupos de doenças etc (44, 48,49). Novas classificações de doenças estão emergindo baseadas na análise do perfil transcricional de uma grande quantidade de genes e, futuramente, de todo o genoma. A grande massa de dados que resulta deste tipo de análise pode também proporcionar uma perspectiva única para o desenvolvimento de terapêuticas baseadas em uma compreensão global da transcrição de certos grupos de genes associados com específicas regiões de sinalização intracelular.

Microarrays de cDNA se baseiam na hibridação complementar de DNA a cDNA. O DNA em questão é imobilizado a uma matriz representando a fase sólida, que é então hibridada a amostras de cDNA a serem testadas (50). Milhares de segmentos de DNA compõem versões modernas de chips e se aproximam da totalidade dos genes existentes nos genomas de várias espécies.

Existem atualmente diversas tecnologias que variam na forma e tamanho do fragmento de DNA que é afixado à fase sólida. Os dois principais grupos

são os *arrays* que utilizam clones (inteiros ou parciais) variando em geral de 500bp a 2 ou 3kb em tamanho e são conhecidos como *arrays* de cDNA. Estes chips são mais flexíveis na sua confecção e economicamente mais vantajosos, e podem conter desde algumas centenas de genes selecionados até cerca de 30.000 genes transcritos não caracterizados representativos do genoma. Usualmente estes chips são hibridados simultaneamente com duas amostras diferencialmente marcadas (figura 2), de forma similar à técnica de CGH, descrita acima. Os resultados são expressos como uma relação entre o perfil de expressão de genes na amostra teste e uma amostra de referência (50,51).

Uma outra variante deste método utiliza uma tecnologia conhecida como fotolitografia que sintetiza oligonucleotídeos (em geral 20 a 25bp de extensão) correspondentes a milhares de genes na própria matriz. Uma coleção de oligonucleotídeos referentes a distintas porções de uma seqüência gênica é utilizada para representar cada um dos genes da matriz, o que permite um alto grau de acurácia dos resultados, apesar do tamanho reduzido da seqüência de DNA. Amostras marcadas com fluorescência (um marcador apenas, neste caso) são usadas para hibridação e o nível de sinal fluorescente emitido por cada fragmento de DNA hibridizado no chip é posteriormente detectado por um leitor (scanner). Um complexo algoritmo é então aplicado para analisar a especificidade do sinal e normalizar os valores em relação ao *background*. Esta metodologia, desenvolvida por uma companhia de biotecnologia, Affymetrix, tem dominado o mercado de chips nos últimos anos.

À medida que a seqüência do genoma humano é refinada e anotada, novas versões de chips irão conter uma representação mais completa de todos os genes e transcritos existentes. Esta informação pode ser assim aplicada à identificação de novos genes que se apresentam anormalmente expressos em certas condições testadas. Assim, pode-se obter, por exemplo, um amplo perfil transcricional de amostras tumorais em comparação ao correspondente tecido normal. O efeito da expressão excessiva, ou da inibição da expressão de certos genes, ou ainda a ação de certas drogas, podem ser avaliadas em relação ao perfil de transcrição de milhares de genes simultaneamente. Análise dos genes (ou grupos de genes) que são ativados ou inativados durante diferentes fases de desenvolvimento embriológico pode ser obtida por este método (52). O uso de chips tem revolucionado a ciência moderna por permitir um acesso sem precedentes a uma enorme variedade de áreas críticas para o estudo biológico, incluindo identificação de genes ou sinalizações celulares, como medida para avaliação diagnóstica e prognóstica, descoberta de novas drogas, na área de toxicologia, de desenvolvimento e envelhecimento.

Aplicação de chips para a caracterização de novos genes é exemplificada na literatura. A caracterização da base molecular da heterogeneidade clínica de certas neoplasias, como certos tipos de linfoma, pode ser finalmente sugerida pelo achado de novos genes que se associam com subgrupos clínicos específicos (53). Em algumas circunstâncias, novos genes podem ser identificados como um defeito comum, mas previa-

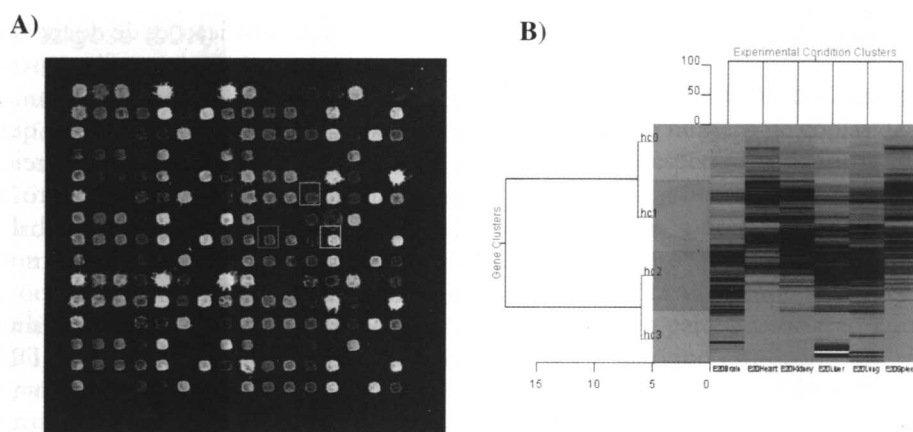


Figura 2. A) Exemplo ilustrativo de um cDNA *array* composto de genes expressos em células precursoras do sistema nervoso central. Amostras de RNA obtido de células tratadas com diferentes fatores de crescimento são marcadas diferencialmente por fluorescência e hibridizadas à matriz. Predomínio da fluorescência de uma das amostras (verde ou vermelho, como indicado) representa uma maior abundância do transcrito respectivo. Quando os transcritos são expressos com intensidade semelhante em ambas as amostras, observa-se uma sobreposição do sinal fluorescente (amarelo). Um dos métodos de análise comparativa de expressão dos genes em um *array* é o dendrograma, que agrupa genes com perfil de expressão semelhante (*hierarchical clustering*). Um exemplo desta análise é apresentado no painel **B**, onde genes (representados por linhas) com alto nível de expressão são representados em vermelho, e genes com baixo nível de expressão, em cor verde, nas diferentes condições experimentais (representadas por colunas).

mente não suspeitado em certos tipos de tumor (54). A caracterização detalhada do papel de tais genes na variedade fenotípica representa um passo importante para o melhor entendimento destas patologias.

Uma outra importante aplicação de chips envolve a caracterização de novos genes responsáveis pela determinação da identidade celular durante a ontogênese de tecidos. Recentemente, identificamos uma série de novos genes associados ao desenvolvimento de linhagens neuronais e astrocíticas a partir do uso combinado de culturas primárias de células primitivas do sistema nervoso central e do uso de uma pré-seleção de clones para análise por cDNA *array* (manuscrito em preparação). Na área de endocrinologia, importantes aplicações de chips incluem a caracterização de diversos tumores cuja base molecular ainda permanece obscura, ou na identificação de fatores moduladores do fenótipo de doenças cuja base genética já é conhecida. Recentemente, potenciais novos genes envolvidos na determinação de ritmos circadianos foram identificados através da análise global de transcrição (55).

Métodos de Clonagem

Uma vez que a área suspeita de conter o gene relacionado à doença tenha sido identificada, o desafio passa então a ser a identificação de genes contidos nesta região para finalmente se caracterizar a natureza do gene alvo. Genes e unidades de transcrição em geral correspondem a apenas 1 a 5% da extensão do DNA genômico, o que pode representar um problema significativo (56). Com o progresso realizado nos últimos meses em relação à anotação do genoma humano (que, a propósito, continua mantendo uma intensa atividade com atualizações diárias) é possível se construir um mapa com alta acurácia e precisão, ao menos no que se refere a genes conhecidos ou experimentalmente comprovados como sendo transcricionalmente ativos (57,58).

Diante de uma longa lista de genes ou transcritos (da ordem de algumas centenas a poucos milhares de genes, a depender da extensão da área e sua densidade gênica), é possível se criar uma estratégia de rastreamento baseada em certos critérios pressupostos ou de fato conhecidos sobre o gene candidato, que possam facilitar a sua identificação. O perfil transcricional dos genes localizados na área candidata pode auxiliar na exclusão (ou, a depender da doença em estudo, inclusão) de genes a serem triados. Estratégias para seqüenciamento em larga escala, utilizando-se bibliotecas de DNA genômico ou cDNA estão entre as opções a serem empregadas nesta fase do processo de identificação do gene. Alternativamente, métodos de

rastreamento utilizando-se genotipagem de alta-densidade (por exemplo, com *arrays* de polimorfismos, descritos abaixo) podem ser empregados como forma de reduzir a área candidata à menor extensão possível através da identificação de novos recombinantes, no caso de desordens hereditárias. Uma outra estratégia envolve a análise do perfil transcricional de genes da região utilizando tecido anormal (por exemplo, tumor) em comparação com uma amostra normal do mesmo tecido em busca de anormalidades representadas por um aumento ou redução da expressão de um ou um grupo de genes como potenciais indicadores de genes a serem rastreados. Um exemplo recente do sucesso desta estratégia de análise transcricional foi utilizada para rastreamento do gene responsável pela susceptibilidade a leiomiomas uterinos e carcinoma renal familiar (59). Embora esta estratégia possa ser importante para acelerar o processo de identificação do gene em questão, uma nota de precaução é necessária, contudo: mutações de genes candidatos nem sempre se acompanham de alterações transcricionais significativas - genes desta natureza, portanto, não poderiam ser detectados se o rastreamento transcricional fosse o único método a ser empregado na procura do gene alvo.

Muitas das técnicas consideradas tradicionais para a clonagem do gene alvo há apenas alguns anos, como o método extensamente usado de *exon trapping* (60), ou técnicas comparativas baseadas em estudos de homólogos de outras espécies, estão atualmente sendo relegadas ao uso em circunstâncias específicas e cada vez menos comuns, em virtude da crescente adoção das mais modernas estratégias descritas acima (46). Contudo, como mencionado, o nível de informação proporcionado pelos bancos de dados disponíveis, em particular os recursos do Centro Nacional de Informação Biotecnológica dos Estados Unidos (NCBI, <http://www.ncbi.nih.gov>) ainda não é completo e algumas áreas do genoma ainda se encontram escassamente representadas em relação à acurácia desta informação. Nestes casos, o uso de estratégias convencionais ainda se faz necessário. Todavia, a incorporação de dados bioinformáticos para a análise de genes e seqüências alcançou tais proporções na prática corrente na área de biologia molecular, que atualmente se considera a identificação de genes por esta via, cunhada de clonagem *in silico*, uma das estratégias mais bem sucedidas no processo da caracterização de genes-alvo. Certamente, a comprovação experimental do achado é um requisito fundamental para o estabelecimento da relação entre um dado gene e o fenótipo em estudo, mas com o uso sis-

temático dos recursos bioinformáticos disponíveis, este processo é consideravelmente simplificado.

Com um certo grau de segurança, é possível especular que, uma vez que todos os genes humanos tiverem sido reconhecidos e mapeados, estratégias de análise de genes candidatos, muitas vezes no passado considerada um exercício de adivinhação (ou, como muitos denominavam, “*a fishing expedition*”), serão o único recurso a ser empregado quando a associação de um gene com um fenótipo particular seja procurada. Haverá uma vasta gama de informações disponíveis para simplificar esta tarefa, como o perfil transcricional, o conhecimento de domínios estruturais da proteína codificada que permitam inferências em relação a função, genes ou produtos associados.

Estudo de polimorfismos e do proteoma: áreas emergentes na ciência molecular

Um importante aspecto da biologia moderna, e que gradualmente tem sido enfatizado, se refere a variações entre indivíduos e seu potencial impacto para o desenvolvimento das mais diversas patologias.

Estas variações, conhecidas como polimorfismos, são freqüentes ao longo do genoma. Estima-se que haja próximo de 2 milhões de polimorfismos de um único nucleotídeo (*single-nucleotide polymorphism* - SNP) distribuídos pelo genoma, dos quais a maior parte se localiza em regiões não codificadoras e apenas uma minoria resulta em modificação do aminoácido codificado. Existem atualmente diversos estudos e programas multicêntricos envolvidos na caracterização da natureza das variações polimórficas de genes humanos, com a intenção de construir um “catálogo” abrangente de tais polimorfismos para estudos de associação. O avanço tecnológico na área de “chips” permite atualmente a análise simultânea de milhares de SNPs e as aplicações desta metodologia crescem a uma velocidade avassaladora (61). Estas plataformas miniaturizadas podem ser empregadas para a caracterização de perfis populacionais, para estudos epidemiológicos, correlações genotípicas com doenças específicas, utilização comparativa em tecidos anormais e normais para análise global de defeitos consistentes, como, por exemplo, áreas de deleções em tumores, associações entre o perfil genotípico e a resposta terapêutica, entre outros (61-63).

A função de polimorfismos na modulação de fenótipos, que se estende muito além do seu tradicional papel de “marcador” genômico, foi recentemente caracterizada em uma descrição de uma região promotora do gene PIG3, cuja seqüência representa um clássico marcador dinucleotídico do tipo “CA” (64). Foi-se observado que variações da extensão deste

marcador afetavam a eficiência da sua resposta promotora. Neste caso, o promotor é responsivo ao supressor de tumor p53. Este achado tem enormes implicações pois, se confirmado em análises de maiores séries, pode indicar que a resposta supressora de tumor de alguns indivíduos pode ser prevista com base no tamanho do alelo associado ao polimorfismo do promotor. É possível que este seja apenas o primeiro exemplo de outras funções moduladoras de fenótipo exercidas por polimorfismos.

Contudo, as variações da seqüência genômica, embora relevantes, quase que certamente não são o único fator na diversidade fenotípica. A integração deste conhecimento com o perfil transcricional de múltiplos genes oferece valiosa informação para associações com situações patológicas específicas.

Um terceiro nível de análise em larga escala além do perfil polimórfico e transcricional está se tornando conhecido como o estudo do proteoma (65). Esta sub-área emergente na ciência moderna engloba a identificação em massa de proteínas e seu papel em regulações pós-tradução; na possibilidade de análises comparativas do perfil proteico global em distintas situações fisiológicas e patológicas de maneira semelhante ao realizado atualmente em transcrição, e também no entendimento de relações interproteicas e genes-proteínas, ou “redes regulatórias”, que definem de forma global as múltiplas interações que fazem um complexo organismo como o humano operar de forma balanceada.

CONCLUSÕES

O número de técnicas e estratégias existentes para a de novos genes é bastante extenso, como visto acima. Os últimos anos têm sido testemunha de uma crescente tendência à adoção de iniciativas multidisciplinares como forma de obter uma melhor compreensão e permitir a exploração deste abundante volume de informação disponível. Tanto os métodos variam em complexidade como também variam as preferências e escolha das estratégias para cada condição particular. A identificação do gene responsável por doenças específicas na verdade é o primeiro passo no longo processo de entendimento dos mecanismos regulatórios que regem o nosso organismo e a evolução humana através dos tempos. Novos modelos e paradigmas biológicos estão sendo propostos e construídos a cada dia, impostos pela pressão desta imensa massa de conhecimento que tem sido gerada. Espera-se que esta informação seja por fim traduzida no desenvolvimento de novas drogas que irão proporcionar o avanço no tratamento de doenças humanas.

AGRADECIMENTOS

Gostaria de agradecer a contribuição do Dr. Ricardo Aguiar e aos Drs. Charles Stiles, John Alberta, Claire Sauvageot e John Park pela cessão de dados pré-publicação para ilustração da figura 2.

REFERÊNCIAS

1. Fearon ER, Cho KR, Nigro JM, Kern SE, Simons JW, Rupert JM, , , et al. Identification of a chromosome 18q gene that is altered in colorectal cancers. **Science** **1990**;247:49-56.
2. Monaco AP, Neve RL, Colletti-Feener C, Bertelson CJ, Kurnit DM, Kunkel LM. Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. **Nature** **1986**;323:646-50.
3. Mulligan LM, Kwok JBJ, Healey CS, Elsdon MJ, Eng C, Gardner E, , , et al. Germline mutations of the *RET* proto-oncogene in multiple endocrine neoplasia type 2A. **Nature** **1993**;363:458-60.
4. Chandrasekharappa SC, Guru SC, Manickam P, Olufemi SE, Collins FS, Emmert-Buck MR, , , et al. Positional cloning of the gene for multiple endocrine neoplasia type 1. **Science** **1997**;276:404-7.
5. Latif F, Tory K, Gnarr J, Yao M, Duh F-M, Orcutt M-L, et al. Identification of the von Hippel-Lindau disease tumor suppressor gene. **Science** **1993**;260:1317-20.
6. Liaw D, Marsh DJ, Li J, Dahia PL, Wang SI, Zheng Z, , et al. Germline mutations of the *PTEN* gene in Cowden disease, an inherited breast and thyroid cancer syndrome. **Nat Genet** **1997**;16:64-7.
7. Kirschner LS, Carney JA, Pack SD, Taymans SE, Giatzakis C, Cho YS, et al. Mutations of the gene encoding the protein kinase A type I-alpha regulatory subunit in patients with the Carney complex. **Nat Genet** **2000**;26:89-92.
8. Larsson C, Skogseid B, Oberg K, Nakamura Y, Nordenskjold M. Multiple endocrine neoplasia type 1 gene maps to chromosome 11 and is lost in insulinoma. **Nature** **1988**;332:85-7.
9. Mathew CGP, Chin KS, Easton DF, Thorpe K, Carter C, Liou GI, et al. A linked genetic marker for multiple endocrine neoplasia type 2A on chromosome 10. **Nature** **1987**;328:527-8.
10. Nelen MR, van Staveren WC, Peeters EA, Hassel MB, Gorlin RJ, Hamm H, et al. Germline mutations in the *PTEN/MMAC1* gene in patients with Cowden disease. **Hum Mol Genet** **1997**;6:1383-7.
11. Marsh DJ, Learoyd DL, Andrew SD, Krishnan L, Pojer R, Richardson A-L, et al. Somatic mutations in the *RET* proto-oncogene in sporadic medullary thyroid carcinoma. **Clin Endocrinol** **1996**;44:249-57.
12. Morton NE. Sequential tests for the detection of linkage. **Am J Hum Genet** **1955**;7:277-318.
13. Ott J. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. **Am J Hum Genet** **1974**;26:588-97.
14. Lathrop GM, Lalouel JM. Easy calculations of lod scores and genetic risks on small computers. **Am J Hum Genet** **1984**;36:460-5.
15. Lathrop GM, Lalouel J-M, Julier C, Ott J. Strategies for multilocus linkage analysis in humans. **Proc Natl Acad Sci USA** **1984**;81:3443-6.
16. Larsson C, Shepherd J, Nakamura Y, Blomberg C, Weber G, Werelius B, et al. Predictive testing for multiple endocrine neoplasia type 1 using DNA polymorphisms. **J Clin Invest** **1992**;89:1344-9.
17. Risch N, Merikangas K. The future of genetic studies of complex human diseases. **Science** **1996**;273:1516-7.
18. Briscoe D, Stephens JC, O'Brien SJ. Linkage disequilibrium in admixed populations: applications in gene mapping. **J Hered** **1994**;85:59-63.
19. Lander ES, Botstein D. Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms. **Proc Natl Acad Sci USA** **1986**;83:7353-7.
20. Call KM, Claser T, Ito CY, Buckler AJ, Pelletier J, Haber DA, et al. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. **Cell** **1990**;60:509-20.
21. Solomon E, Voss R, Hall V, Bodmer WF, Jass JR, Jeffreys AJ, et al. Chromosome 5 allele loss in human colorectal carcinomas. **Nature** **1987**;328:616-9.
22. Nath J, Johnson KL. A review of fluorescence in situ hybridization (FISH): current status and future prospects. **Biotech Histochem** **2000**;75:54-78.
23. Chang SS, Mark HF. Emerging molecular cytogenetic technologies. **Cytobios** **1997**;90:7-22.
24. Forozan F, Karhu R, Kononen J, Kallioniemi A, Kallioniemi OP. Genome screening by comparative genomic hybridization. **Trends Genet** **1997**;13:405-9.
25. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. **Proc Natl Acad Sci USA** **1971**;68:820-3.
26. Knudson AG Jr. Overview: genes that predispose to cancer. **Mutat Res** **1991**;247:185-90.
27. Kuska B. Alfred Knudson: two hits times 25 years. **J Natl Cancer Inst** **1997**;89:470-3.
28. Hansen MF, Cavenee WK. Genetics of cancer predisposition. **Cancer Res** **1987**;47:5518-27.
29. Khosla S, Patel VM, Hay ID, Schaid DJ, Grant CS, van Heerden JA, et al. Loss of heterozygosity suggests multiple genetic alterations in pheochromocytomas and medullary thyroid carcinomas. **J Clin Invest** **1991**;87:1691-9.
30. Dong Q, Debelenko LV, Chandrasekharappa SC, Emmert-Buck MR, Zhuang Z, Guru SC, et al. Loss of heterozygosity at 11q13: analysis of pituitary tumors, lung carcinoids, lipomas, and other uncommon tumors in subjects with familial multiple endocrine neoplasia type 1. **J Clin Endocrinol Metab** **1997**;82:1416-20.
31. Farrell WE, Simpson DJ, Bicknell JE, Talbot AJ, Bates AS, Clayton RN. Chromosome 9p deletions in invasive and noninvasive nonfunctional pituitary adenomas: the deleted region involves markers outside of the *MTS1* and *MTS2* genes. **Cancer Res** **1997**;57:2703-9.
32. Yamada S, Yoshimoto K, Sano T, Takada K, Itakura M, Usui M, et al. Inactivation of the tumor suppressor gene

- on 11q13 in brothers with familial acrogigantism without multiple endocrine neoplasia type 1. **J Clin Endocrinol Metab** 1997;82:239-42.
33. Bose S, Wang SI, Terry MB, Hibshoosh H, Parsons R. Allelic loss of chromosome 10q23 is associated with tumor progression in breast carcinomas. **Oncogene** 1998;17:123-7.
34. Gilbert F. Medical genetics. Retinoblastoma and recessive alleles in tumorigenesis. **Nature** 1983;305:761-2.
35. Yandell DW, Poremba C. Genetics of retinoblastoma: implications for other human cancers. **Med Pediatr Oncol Suppl** 1996;1:25-8.
36. Orkin SH. Reverse genetics and human disease. **Cell** 1986;47:845-50.
37. Zhuang Z, Bertheau P, Emmert-Buck MR, Liotta LA, Gnarr J, Linehan WM, et al. A microdissection technique for archival DNA analysis of specific cell populations in lesions < 1mm in size. **Am J Pathol** 1995;146:620-5.
38. Houldsworth J, Chaganti RS. Comparative genomic hybridization: an overview. **Am J Pathol** 1994;145:1253-60.
39. Pinkel D, Seagraves R, Sudar D, Clark S, Poole I, Kowbel D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. **Nat Genet** 1998;20:207-11.
40. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. **Nat Genet** 1999;23:41-6.
41. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. **Science** 1992;257:967-71.
42. Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. **Science** 1993;259:946-51.
43. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. **Science** 1995;270:484-7.
44. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. **Science** 1999;286:531-7.
45. Dahia PLM, Eng C. Genetic Disorder of Endocrine Neoplasia. 1 ed. *Frontiers in Hormone Research*, ed. Grossman AB. Vol. 28. 2001, Basel: Karger. 220.
46. Aguiar RCT, Dahia PLM. Identification and characterization of disease-related genes: focus on endocrine neoplasias, in *Genetic Disorder of Endocrine Neoplasia*, Dahia PLM and Eng C, Editors. 2001, Karger: Basel. p. 20-49.
47. Lockhart DJ, Winzler EA. Genomics, gene expression and DNA arrays. **Nature** 2000;405:827-36.
48. Cho RJ, Campbell MJ, Winzler EA, Steinmetz L, Conway A, Wodicka L, et al. A genome-wide transcriptional analysis of the mitotic cell cycle. **Mol Cell** 1998;2:65-73.
49. Alizadeh AA, Staudt LM. Genomic-scale gene expression profiling of normal and malignant immune cells. **Curr Opin Immunol** 2000;12:219-25.
50. Bowtell DD. Options available — from start to finish — for obtaining expression data by microarray. **Nat Genet** 1999;21:25-32.
51. Cheung VG, Morley M, Aguilar F, Massimi A, Kucherlapati R, Childs G. Making and reading microarrays. **Nat Genet** 1999;21:15-9.
52. Schena M, Heller RA, Theriault TP, Konrad K, Lachmeier E, Davis RW. Microarrays: biotechnology's discovery platform for functional genomics. **Trends Biotechnol** 1998;16:301-6.
53. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RCT, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. **Nat Med** 2002;8:68-74.
54. Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. **Nat Genet** 2002;30:41-7.
55. Storch K, Lipan O, Leykin I, Viswanathan N, Davis F, Wong W, et al. Extensive and divergent circadian gene expression in liver and heart. **Nature** 2002;417:78-83.
56. Lewin B. **Genes V**, ed. Lewin B. 1994, Oxford: Oxford University Press. 1272.
57. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. **Nature** 2001;409:860-921.
58. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. **Science** 2001;291:1304-51.
59. Tomlinson IP, Alam NA, Rowan AJ, Barclay E, Jaeger EE, Kelsell D, et al. Germline mutations in FH predispose to dominantly inherited uterine fibroids, skin leiomyomata and papillary renal cell cancer. **Nat Genet** 2002;30:406-10.
60. Duyk GM, Kim SW, Myers RM, Cox DR. Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. **Proc Natl Acad Sci USA** 1990;87:8995-9.
61. Chanock S. Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. **Dis Markers** 2001;17:89-98.
62. Miller RD, Kwok PY. The birth and death of human single-nucleotide polymorphisms: new experimental evidence and implications for human history and medicine. **Hum Mol Genet** 2001;10:2195-8.
63. Nowotny P, Kwon JMAMG. SNP analysis to dissect human traits. **Curr Opin Neurobiol** 2001;11:637-41.
64. Contente A, Dittmer A, Koch MC, Roth JMD. A polymorphic microsatellite that mediates induction of PIG3 by p53. **Nat Genet** 2002;30:315-20.
65. Pandey A, Mann M. Proteomics to study genes and genomes. **Nature** 2000;405:837-46.

Endereço para correspondência:

Patricia Dahia
44 Binney Street SM1010
02115-6084 Boston, MA
Fax: (617) 632-4663
e.mail: patricia_dahia@dfci.harvard.edu