

Educacional / Educational

Para que existem as regras de nomenclatura genética? *Why are there rules for genetic nomenclature?*

Alessandra Splendore

Na comunicação científica, precisamos respeitar os princípios de clareza, precisão, comunicabilidade e consistência. Para isso, existe um comitê de nomenclatura genética, cuja função é assegurar que cada gene humano tenha um nome e símbolo únicos que sejam usados consistentemente na literatura científica. Apesar dos esforços, ainda encontramos textos onde o autor se refere a um gene usando um símbolo obsoleto, ou não faz a distinção adequada entre o gene e a proteína, prejudicando a compreensão por parte do leitor. Com isso em mente, apresento uma revisão das regras de nomenclatura de genes e diversos recursos disponíveis através da Internet para encontrar os nomes adequados dos genes, enfatizando aqueles envolvidos em câncer. Rev. bras. hematol. hemoter. 2005;27(2):148-152.

Palavras-chave: Nomenclatura; translocação; genética; classificação.

Introdução

Na comunicação científica, é de fundamental importância o uso de termos precisos que transmitam de forma clara e concisa as idéias dos autores aos leitores. Para isso, existem convenções de nomenclatura que, quando conhecidas e empregadas por todos, facilitam o entendimento entre os profissionais da área. A exemplo do que ocorre na Medicina, que adota os termos reunidos na *Nomina Anatomica* para designar as estruturas do corpo, também na Genética existem normas para a nomenclatura de genes, embora estas nem sempre sejam respeitadas. Os problemas mais freqüentes relacionados à nomenclatura de genes em trabalhos científicos são autores que se referem ao mesmo gene usando nomes diferentes e autores que não deixam claro se estão se referindo ao gene ou à proteína.

A existência de problemas na nomenclatura de genes foi percebida já na década de 1960, quando ainda existiam poucos genes humanos identificados, e as primeiras normas de nomenclatura foram apresentadas em 1979.^{1,2} Na última década, com o avanço das técnicas de mapeamento e clonagem gênica e a conclusão do Projeto Genoma Humano, o número de genes humanos conhecidos aumentou exponencialmente, em um espaço de tempo relativamente

curto. O grande número de genes identificados e arquivados em diversos bancos de dados eletrônicos torna ainda mais urgente o estabelecimento de uma nomenclatura universal e única para todos os genes. A entidade que assegura que cada gene tenha um nome e um símbolo únicos é um subcomitê do "The Human Genome Organization" (HUGO),³ órgão internacional dedicado à coordenação de esforços colaborativos na pesquisa sobre o genoma humano. O "HUGO Gene Nomenclature Committee" (HGNC)^{1,4} tem a função de aprovar um nome e um símbolo para cada gene catalogado, seguindo certas regras que visam facilitar tanto a comunicação entre pesquisadores como o armazenamento e busca em bancos de dados eletrônicos. Cada gene recebe um nome que deve ser curto e trazer informações sobre sua função e sua relação com outros genes da mesma família. Muitas vezes, nomes antigos ou de síndromes associadas a defeitos no gene são incorporadas ao nome oficial do gene, entre parênteses [exemplo: tumor protein p53 (Li-Fraumeni syndrome); cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)]. Além do nome oficial, cada gene tem um símbolo, uma notação curta usada para designar o gene em bancos de dados e publicações. Os principais bancos de dados que armazenam informações sobre genes humanos utilizam os símbo-

Divisão de Genética, Instituto Nacional de Câncer, Rio de Janeiro-RJ.

Endereço para correspondência: Divisão de Genética – Instituto Nacional de Câncer – INCa-RJ.
Rua André Cavalcanti, 37, 4º andar
20231-050 Centro – Rio de Janeiro-RJ
Tel.: 3233-1486 ou 3233-1460 - Fax: 2226-0875
E-mail: ale.splendore@terra.com.br

los aprovados pelo HGNC, o que assegura que a troca de dados entre os diferentes bancos possa ser feita de modo inequívoco. O uso dos símbolos deve obedecer a algumas regras de grafia, listadas na Tabela 1.

Já existem mais de 20 mil genes catalogados no banco de dados do HGNC, chamado *Genew*,⁵⁻⁷ que, além do nome

Regras	Exemplos	
	Certo	Errado
Genes humanos devem ser grafados sempre em maiúsculas e em itálico	<i>BRCA1</i>	BRCA1 <i>Brca1</i>
Nos símbolos não são permitidos hífens, sobre ou subscritos, algarismos romanos ou letras do alfabeto grego	<i>IGF2</i> <i>TGFB3</i>	<i>IGF-II</i> TGFβ3
Não se deve usar "g" (de gene) ou "h" (de humano) antes do símbolo	<i>MSH2</i>	<i>hMSH2</i>
Não se deve usar o prefixo "c-" para designar oncogenes celulares	<i>MYC</i> <i>HRAS</i>	<i>c-myc</i> c-H-ras

Core Data		Database Links	
Approved Symbol +	TP53	Accession Numbers +	
Approved Name +	tumor protein p53 (Li-Fraumeni syndrome)	AF307851	GenBank
HGNC ID +	11998	Pubmed IDs +	
Status +	Approved	2047879	PMID
Chromosome +	17p13.1	OMIM ID (mapped data) +	
Previous Symbols +		191170	OMIM
Previous Names +		LocusLink ID (mapped data) +	
Aliases +	P53	7157	Gene Map Viewer
		RefSeq (mapped data) +	
Gene Symbol Links		NM_000546	GenBank UCSC
Ensembl GeneView GENATLAS GeneCards GeneClinics/GeneTests		UniProt ID (mapped data) +	
		P04637	SwissProt UniProt

Figura 1: Entrada do banco de dados *Genew* sobre o gene *TP53*, contendo informações essenciais e links para outros bancos de dados

Approved Gene Symbol	Approved Gene Name	Location	Aliases
DDEF1	Development and differentiation enhancing factor 1	8q24.1-Q24.2	ASAP1, PAP , ZG14P
DDEF2	development and differentiation enhancing factor 2	2p	PAP , SHAG1
MRPS30	mitochondrial ribosomal protein S30	5q11	PAP , PDCD9
PAPOLA	poly(A) polymerase alpha	14q32.1-q32.2	PAP
PDAP1	PDGFA associated protein 1	7q	PAP1 , PAP , HASPP28
TUSC2	tumor suppressor candidate 2	3p21.3	FUS1 , PAP

Figura 2. Resultado da busca do banco de dados *Genew* por genes cujo codinome (*alias*) é "PAP"

e símbolo oficiais, informa também a localização cromossômica e os codinomes (*aliases*, outros símbolos historicamente associados ao mesmo gene) mais frequentes. Além dessas informações, o *Genew* fornece links para diversos bancos de dados, como OMIM,⁸ GenBank,⁹ Ensembl¹⁰ e UniProt,¹¹ entre outros, como pode ser visto na Figura 1. O HGNC recomenda que os nomes/símbolos para designar proteínas sejam os mesmos usados para os genes, embora existam inúmeras exceções. Existem comitês próprios para nomenclatura de enzimas e proteínas de classes específicas e, nesses casos, a nomenclatura dos genes deve respeitar os nomes já estabelecidos das proteínas (exemplo: hemoglobina, alpha 1; símbolo: *HBA1*). Quando o mesmo símbolo é usado para o gene e a proteína, a maneira de diferenciá-los é pelo uso de itálico (ex: gene *GATA1*, proteína GATA-1). O uso consistente dessa regra simples é suficiente para esclarecer ao leitor se o autor do trabalho está se referindo ao gene ou à proteína.²

Outra distinção importante é feita entre a nomenclatura de genes humanos e seus homólogos no camundongo.

Como o camundongo é o modelo preferido para o estudo de alterações genéticas relevantes para a saúde humana, existe um esforço para manter um paralelo entre o sistema de nomenclatura de genes humanos e murinos. Para isso, o órgão responsável pela nomenclatura de genes do camundongo, o "International Committee on Standardized Genetic Nomenclature for Mice",¹² trabalha em parceria com o HGNC e, assim, para a grande maioria dos genes, o mesmo símbolo é utilizado na designação do gene humano e o seu homólogo murino. Os símbolos dos genes do homem e do camundongo são diferenciados pelo uso de maiúsculas: os genes murinos devem ser grafados em itálico, com a primeira letra maiúscula seguida de letras minúsculas. Assim, *BRCA1* é o gene humano, enquanto *Brca1* é o gene homólogo do camundongo. Como é freqüente encontrar em uma mesma publicação referência tanto ao gene humano como murino, a diferença de grafia entre ambos é o que esclarece ao leitor a que organismo pertence o gene em questão.

Genes que mudam de nome

Um dos problemas enfrentados pelo HGNC ao aprovar um nome e símbolo únicos para cada gene foi a existência de genes que haviam sido descritos por mais de um autor e recebido nomes diferentes. Por exemplo, em 1998, cinco grupos diferentes descreveram um

Tabela 2
Símbolos e nomes de alguns dos genes importantes no estudo do câncer

Símbolo	Nome	Alias
<i>TP53</i>	tumor protein p53 (Li-Fraumeni syndrome)	P53
<i>RB1</i>	retinoblastoma 1 (including osteosarcoma)	RB
<i>ERBB2</i>	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	NEU, HER-2
<i>MYC</i>	v-myc myelocytomatosis viral oncogene homolog (avian)	c-Myc
<i>MYCN</i>	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)	n-Myc, NMYC
<i>HRAS</i>	v-Ha-ras Harvey rat sarcoma viral oncogene homolog	c-H-ras, RASH1, p21
<i>NRAS</i>	v-Ha-ras Harvey rat sarcoma viral oncogene homolog	c-N-ras, N-Ras, p21
<i>KRAS2</i>	v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog	c-K-ras, KI-RAS, RASK2, p21
<i>CDKN1A</i>	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	p21, CIP1, WAF1
<i>CDKN2A</i>	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	p16, p14, INKa, ARF
<i>MLH1</i>	mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)	hMLH1, HNPCC
<i>MSH2</i>	mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)	hMSH2, HNPCC
<i>RUNX1</i>	runt-related transcription factor 1 (acute myeloid leukemia 1; aml1 oncogene)	AML1, CBFA2
<i>RUNX1T1</i>	runt-related transcription factor 1; translocated to, 1 (cyclin D-related)	AML1T1, CBF2T1, CDR, MTG8, ETO
<i>ETV6</i>	ets variant gene 6 (TEL oncogene)	TEL
<i>MLL</i>	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)	TRX1, ALL-1, HRX, HTRX1
<i>MLL2</i>	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 2	AF-4
<i>MLL3</i>	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 3	AF-9, AF9
<i>TCF3</i>	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)	E2A, ITF1

mesmo gene no cromossomo 11, possivelmente relacionado ao desenvolvimento tumoral, que foi apresentado na literatura como *ITM*,¹³ *BWR1A*,¹⁴ *IMPT1*,¹⁵ *TSSC5*¹⁶ ou *ORCTL2*.¹⁷ Por fim, o gene teve como símbolo aprovado uma sexta sigla, *SLC22A1L*. Além da existência de genes com diversos nomes, também há casos em que um mesmo símbolo foi associado a genes diferentes. O exemplo mostrado na Figura 2 revela que o símbolo *PAP* já foi usado para designar seis genes distintos. Para evitar a troca excessiva de nomes, muitas revistas passaram a exigir que os autores obtenham um símbolo aprovado pelo HGNC antes de publicar a descrição de um novo gene.

Uma vez aprovado um nome e um símbolo para um gene, ainda resta o problema de os nomes alternativos permanecerem em uso na literatura. Mesmo no caso de genes que não receberam mais de um nome, o símbolo aprovado pelo HGNC muitas vezes difere daquele pelo qual o gene se tornou primariamente conhecido na literatura. Isso se aplica particularmente aos oncogenes, que figuram entre os primeiros genes humanos identificados. Historicamente, proto-oncogenes (ou oncogenes celulares) eram descritos pela designação do gene viral com o qual apresentavam homologia e o acréscimo do prefixo "c-" para designar "celular" ou

"constitutivo" e assim diferenciá-los dos oncogenes virais. Desse modo, os primeiros proto-oncogenes foram designados c-myc, c-H-ras, c-Ki-Ras, c-Kit e assim por diante. Para que todos os símbolos de genes sigam as mesmas regras de grafia e também para facilitar a procura por símbolo nos bancos de dados (onde a presença de hífen é detrimental), esses genes tiveram os nomes atualizados (*MYC*, *HRAS*, *KRAS2*, *KIT*). Contudo, muitos autores se acostumam a usar um determinado símbolo consagrado pela literatura e relutam em adotar os símbolos aprovados pelo HGNC. Essa resistência é particularmente sentida em relação aos genes que se encontram envolvidos em translocações cromossômicas nas leucemias, pois muitas vezes o gene quimérico assume uma identidade própria. Por exemplo, ao ler *TEL-AML1*, o hematologista oncológico reconhece de imediato que se trata da translocação t(12;21) da leucemia linfoblástica aguda, embora a forma correta de descrever o gene quimérico seja *ETV6-RUNX1*.

Alguns exemplos de genes freqüentemente envolvidos em translocações nas leucemias cujo nome aprovado não é o mais utilizado são: *RUNX1* (*AML1*, *CBFA2*), *ETV6* (*TEL*), *RUNX1T1* (*ETO*), *MLL* (*HRX*) e *TCF3* (*E2A*). Mesmo que os autores prefiram utilizar os símbolos antigos, é funda-

mental pelo menos conhecer o nome correto do gene e, mais importante, citá-lo nos trabalhos publicados. Muitas vezes, os símbolos antigos podem não ter registro nos bancos de dados ou encontrar-se associados a mais de um gene. Conhecer o símbolo aprovado de um gene também é extremamente relevante na realização de levantamentos bibliográficos, pois publicações recentes em periódicos de ponta poderão não ser encontrados nas buscas eletrônicas realizadas usando apenas o símbolo antigo. A Tabela 2 lista exemplos dos genes mais citados em trabalhos relacionados à genética do câncer, indicando o símbolo correto do gene, seu nome e os codinomes mais freqüentes.

Gene ou proteína?

Na maioria dos casos, a proteína é designada pelo mesmo símbolo do gene. As notórias exceções são as proteínas dos genes *RBI* e *TP53*, que são comumente designadas pRB e p53. A designação de proteínas pela expressão "p###" surgiu como modo de identificar produtos gênicos descobertos pelo seu papel na regulação do ciclo celular, sem que se tivesse conhecimento sobre os genes que as codificavam. Desse modo, a descoberta de uma proteína de peso molecular igual a 53 quilo Daltons (kDa) lhe valia o nome p53. A designação de um produto gênico por seu peso molecular é extremamente inadequada, pois além de não nos informar sobre a sua função, produtos com o mesmo peso molecular podem ser codificados por genes diferentes. Por exemplo, existem pelo menos 4 genes relevantes para câncer que codificam proteínas de 21 kDa que já foram chamadas de "p21": *CDKN1A* e três genes da família Ras, *HRAS*, *NRAS* e *KRAS2*. Se um determinado autor diz que em seu trabalho estudou o gene (ou a proteína) p21, o leitor não tem como saber qual desses genes foi estudado nem a função da proteína.

Burlando as próprias regras

Um dos motivos que causa confusão é que nem sempre, na literatura, as regras de nomenclatura e grafia são seguidas à risca. Principalmente no meio eletrônico (Internet) é comum encontrarmos símbolos de genes grafados sem itálico. Muitas revistas que oferecem artigos no formato html não usam itálico em seus textos on-line, mas usarão a grafia correta na edição impressa ou na versão portable document file (.pdf). Mais grave do que a grafia correta, entretanto, é o fato de nem todas as revistas pedirem que os autores utilizem os símbolos de genes aprovados para a publicação de artigos. O HGNC não tem como exigir o cumprimento das regras de nomenclatura genética, cabendo-lhe apenas propor um nome e símbolo oficiais para cada gene. Quem detém o poder de fazer com que os autores respeitem as normas de nomenclatura genética são os editores das revistas científicas. O evidente benefício do uso universal de nomes e símbolos consistente para os genes deve fazer com que um número

cada vez maior de revistas passe a exigir o cumprimento dessas regras em suas publicações, a exemplo do que já fazem as principais revistas da área biomédica, como a *Nature*, *Nature Genetics*, *American Journal of Human Genetics*, *Genomics*, *Human Mutation*, *Genes Chromosomes & Cancer* e *The Lancet*.

Onde encontrar os símbolos aprovados pelo HGNC

Para saber qual o símbolo aprovado de um determinado gene, deve-se consultar a página do HGNC/*Genew*.⁷ Na página principal, há um campo para busca (*search*) onde é possível colocar o nome do gene por extenso, seu símbolo ou codinome. Nesse campo, não devem ser incluídos números romanos e prefixos "c" ou "h".

Outra maneira fácil de checar o nome do gene é procurá-lo pelo OMIM⁸ (*Online Mendelian Inheritance in Man*), catálogo de doenças genéticas e genes humanos, onde a busca pode ser realizada pela doença associada ao gene, seu nome ou símbolo. A página de resultados trará um breve histórico e revisão básica da literatura, além de vários links na barra à esquerda do texto. Selecionando Nomenclature entre os links, será aberta a página referente ao gene em questão no *Genew*. Uma boa referência para procurar informações sobre genes envolvidos em translocações nas leucemias é o *Atlas of Genetics and Cytogenetics in Oncology and Haematology*.¹⁸ Nessa página, é possível procurar os genes pelo tipo de translocação em que se encontram envolvidos [por exemplo, t(9;22)(q34;q11)] ou tipo de leucemia. Apesar de a maioria dos genes estarem listados na página com os nomes mais comuns, clicando sobre o nome do gene aparecerá uma série de informações sobre ele, inclusive o nome aprovado pelo HGNC e outros codinomes. Clicando sobre o nome do gene indicado ao lado da designação "HUGO", será aberta a página relativa ao gene no *Genew*.

Abstract

In scientific writings, we must strive for clarity, accuracy, objectivity, and consistency. Hence the purpose of a genetic nomenclature committee that assigns a unique name and symbol for each human gene, ensuring their unequivocal identification in the scientific literature. Despite these efforts, we still see publications using outdated symbols for genes, and authors that fail to distinguish between gene and protein in their writings. Since such mistakes can severely hinder the readers' comprehension, I present a review of gene nomenclature guidelines, and of the various Internet tools for retrieving approved gene names and symbols, highlighting those that are important in the study of cancer. Rev. bras. hematol. hemoter. 2005;27(2):148-152.

Key words: Nomenclature; translocation.

Referências Bibliográficas

1. Povey S, Lovering R, Bruford E, et al. The HUGO Gene Nomenclature Committee (HGNC). *Hum Genet* 2001;109:678-680.
2. Wain HM, Lovering RC, Bruford EA, et al. Guidelines for human gene nomenclature. *Genomics* 2002;79:464-470.
3. The Human Genome Organisation: <http://www.gene.ucl.ac.uk/hugo/>
4. HUGO Gene Nomenclature Committee: <http://www.gene.ucl.ac.uk/nomenclature/>
5. Wain HM, Lush M, Ducluzeau F, et al. the Human Gene Nomenclature Database. *Nucleic Acids Res* 2002;30:169-171.
6. Wain HM, Lush MJ, Ducluzeau F, et al. Genew: the Human Gene Nomenclature Database, 2004 updates. *Nucleic Acids Res* 2004;32 Database issue:D255-257.
7. Genew: <http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl>
8. Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 2000:<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>.
9. GenBank, gene database of the National Center for Biotechnology Information (NCBI): <http://www.ncbi.nlm.nih.gov/>
10. Ensembl: http://www.ensembl.org/Homo_sapiens/
11. UniProt, the universal protein resource: <http://www.pir.uniprot.org/>
12. MGI - Mouse Genome Informatics; Mouse Nomenclature Home Page:<http://www.informatics.jax.org/mgihome//nomen/index.shtml>
13. Morisaki H, Hatada I, Morisaki T, et al. A novel gene, ITM, located between p57KIP2 and IPL, is imprinted in mice. *DNA Res* 1998;5:235-240
14. Schwienbacher C, Sabbioni S, Campi M, et al. Transcriptional map of 170-kb region at chromosome 11p15.5: identification and mutational analysis of the BWR1A gene reveals the presence of mutations in tumor samples. *Proc Natl Acad Sci USA* 1998; 95:3.873-3.878.
15. Dao D, Frank D, Qian N, et al. IMPT1, an imprinted gene similar to polyspecific transporter and multi-drug resistance genes. *Hum Mol Genet* 1998;7:597-608
16. Lee MP, Reeves C, Schmitt A, et al. Somatic mutation of TSSC5, a novel imprinted gene from human chromosome 11p15.5. *Cancer Res* 1998;58:4155-4159.
17. Cooper PR, Smilnich NJ, Day CD, et al. Divergently transcribed overlapping genes expressed in liver and kidney and located in the 11p15.5 imprinted domain. *Genomics* 1998;49:38-51
18. Atlas of Genetics and Cytogenetics in Oncology and Haematology. <http://www.infobiogen.fr/services/chromcancer/>

Avaliação: Editor e dois revisores externos

Conflito de interesse: não declarado

Recebido: 20/04/05

Aceito após modificações: 16/05/05