The International Journal of Biological Markers, Vol. 23 no. 2, pp.

Poor usage of HUGO standard gene nomenclature in cancer marker studies

M. Lacroix

InTextoResearch, Baelen, Wallonia - Belgium

ABSTRACT: Since 1989, the Human Genome Organisation (HUGO) attempts to ensure that for each gene there is one name and one symbol. The resulting standard nomenclature is, however, poorly applied in clinical studies, which impairs the efficient retrieval of information. This lack of support is reflected in the present survey of articles reporting on disseminated tumor cell detection. (Int J Biol Markers 2008; 23:)

Key words: Cancer, Gene, Name, Symbol, Standard nomenclature, Human Genome Organisation, HUGO, Clinical studies, Disseminated tumor cells, RT-PCR

Analysis of gene expression in cancer cells, notably by the reverse-transcriptase polymerase chain reaction (RT-PCR), is widely performed. One promising use of RT-PCR is for the detection of disseminated cancer cells in lymph nodes, peripheral blood and bone marrow (1-3). Many genes of interest in cancer possess multiple names, which may be confusing. However, since 1989, the Human Genome Organisation (HUGO, http://www.hugo-international.org) attempts to ensure that for each gene there is one name and one symbol (4, 5).

To determine to which extent the HUGO nomenclature is used, I recently performed a PubMed search for abstracts mentioning disseminated cancer cell detection by RT-PCR in cancer patients. The terms "disseminated", "circulating", "bone marrow", "blood", or "lymph node(s)" were used as keywords for the search. The resulting 303 articles, covering 15 cancer types (Tab. I) and the years 1999-2007 (Tab. II), were searched for occurrences of standard HUGO gene symbols and names, and of other descriptors ("aliases") for the markers targeted by RT-PCR.

As shown in Table III for 17 markers specific for various cancers, the use of HUGO symbols and names is dramatically low, as compared to that of aliases. SCGB2A2 (gene name: "Secretoglobin family 2A, member 2") provides a particularly striking example. This symbol was not found in 34/34 abstracts of papers reporting on the detection of the gene it designates. Instead, the gene was variably symbolized as MAM, hMAM, hMAM-A, MG, MGB1 or MMG, reflecting at least 3 different names: mammaglobin, mammaglobin A, and mammaglobin 1. Many people performing marker analysis in breast cancer are familiar with the word

TABLE I - ARTICLES MENTIONING DISSEMINATED CANCER CELL DETECTION BY RT-PCR IN CANCER PATIENTS: DISTRIBUTION BY TYPE OF CANCER AND PERIOD OF TIME COVERED BY THE SEARCH

Type of cancer	Number of articles and period covered
Breast	89 articles, between 1999 and 2008
Liver	22 articles, between 2002 and 2007
Lung	19 articles, between 1999 and 2007
Melanoma	34 articles, between 1999 and 2007
Ovary	4 articles, between 2000 and 2006
Prostate	19 articles, between 2002 and 2007
Bladder	14 articles, between 1998 and 2006
Cervical	3 articles, between 1999 and 2007
Colorectal	51 articles, between 2000 and 2007
Esophagus	10 articles, between 2003 and 2007
Gastric	25 articles, between 1999 and 2007
Head & Neck	4 articles, between 1999 and 2007
Pancreas	8 articles, between 1999 and 2007
Kidney	5 articles, between 2003 and 2006
Thyroid	16 articles, between 2000 and 2006

TABLE II - ARTICLES MENTIONING DISSEMINATED CANCER CELL DETECTION BY RT-PCR IN CANCER PATIENTS: DISTRIBUTION BY YEAR

Year	Number of articles
1998	1
1999	17
2000	17
2001	21
2002	36
2003	44
2004	42
2005	50
2006	46
2007	29

0393-6155/031-04$25.00/0

TABLE III - A LIST OF MARKERS USED TO DETECT DISSEMINATED CANCER CELLS FROM SPECIFIC CANCERS BY RT-PCR

HUGO gene symbol (number of abstracts citing this symbol / total number of abstracts citing the marker)	**HUGO gene name (number of abstracts citing this name / total number of abstracts citing the marker)**	**Other frequently-used names and symbols ("aliases")**	**Type of cancer**
ANKRD30A (0 / 5)	Ankyrin repeat domain 30A (0 / 5)	B726; B726P; NY-BR-1	Breast
DCT (0 / 2)	Dopachrome tautomerase (0 / 2)	Tyrosinase-related protein-2; TRP-2	Melanoma
FOLH1 (0 / 6)	Folate hydrolase 1 (0 / 6)	Prostate-specific membrane antigen; PSM; PSMA	Prostate
GUCY2C (0 / 4)	Guanylate cyclase 2C (0 / 4)	Guanylyl cyclase C; guanylylcyclase C; GCC	Colorectal
KLK3 (1 / 14)	Kallikrein-related peptidase 3 (1 / 14)	Prostate specific antigen; prostate-specific antigen; PSA	Prostate
MLANA (0 / 17)	Melan-A (7 / 17)	Melanoma antigen recognized by T cells 1; MART1; MART-1	Melanoma
MUCL1 (0 / 2)	Mucin-like 1 (0 / 2)	Small mucin-like protein; small breast epithelial mucin; BS106; SBEM	Breast
OR51E2 (0 / 2)	Olfactory receptor, family 51, subfamily E, member 2 (0 / 2)	Prostate-specific G protein coupled receptor; prostate-specific gene with homology to G protein receptor; PSGR; STAG1/PMEPA1	Prostate
PLUNC (0 / 2)	Palate, lung and nasal epithelium carcinoma associated (0 / 2)	Lung-specific X-protein; lunx	Lung
SCGB2A1 (0 / 8)	Secretoglobin family 2A, member 1 (0 / 8)	MammaglobinB; mammaglobin-B; mammaglobin 2; mamB; MGB2	Breast
SCGB2A2 (0 / 34)	Secretoglobin family 2A, member 2 (0 / 34)	Mammaglobin; mammaglobin A; mammaglobin 1;MAM; mam; hMAM; hMAM-A; MG; MGB1; MMG	Breast
SERPINB3 (0 / 4)	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 3 (0 / 4)	Squamous cell carcinoma antigen 1; SCCA1; SCC	Esophagus and Head & Neck
SERPINB4 (0 / 1)	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 4 (0 / 1)	Squamous cell carcinoma antigen 2 (SCCA2)	Esophagus and Head & Neck
SERPINB5 (0 / 12)	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 5 (0 / 12)	Maspin; MAS	Breast
SILV (0 / 3)	Silver homolog (mouse) (0 / 3)	Gp100, GP-100	Melanoma
TFF1 (1 / 4)	Trefoil factor 1 (1 / 4)	PS2	Breast
UCHL1 (0 / 2)	Ubiquitin carboxyl-terminal esterase L1 (0 / 2)	Ubiquitin thiolesterase; PGP 9.5	Lung

"mammaglobin". However, there are 2 different mammaglobins, and a PubMed search using the word "mammaglobin" does not retrieve all papers reporting on SCGB2A2 expression analysis.

Another example of an "unpopular" gene symbol is KLK3, the HUGO symbol for the gene otherwise known as PSA (prostate-specific antigen). A PubMed search using KLK3 (without any other restriction) retrieves a total of only 53 articles published between 1994 and May 2008 (Tab. IV). For the same period, nearly 10,000 articles may be found using PSA as keyword. However, among these are only 27 of the 53 papers identified with KLK3. Thus, the use of multiple names for the same gene impairs the efficient retrieval of information from databases.

These observations support those of Tamames and Valencia (6). These authors analyzed PubMed abstracts for the period 1994-2004, collecting information regarding the mention of human gene symbols and the frequency with which official symbols were mentioned in comparison with their aliases. From their study, it was obvious that the scientific community had not widely adopted the HUGO guidelines. Moreover, they observed that the proportion of official symbols that were used predominantly had only increased slightly, from 35% in 1994 to 44% in 2004. Thus, aliases were used far more often than official symbols. Examination of the data on a year-by-year basis indicated that the tendency to improve the situation by replacing aliases in favor of official HUGO symbols was weak. The changes in name usage, either from official to aliases or from aliases to official, were not very frequent, and the nomenclature of most genes remained rather stable with time.

TABLE IV - IMPAIRED RETRIEVAL OF ARTICLES ACCORDING TO THE GENE NAME USED

Year	Articles found using "KLK3" (*) as keyword	Articles from (*) also found using "PSA" as keyword	Percentage
1994	1	1	100%
1995	0	0	-
1996	2	1	50%
1997	2	1	50%
1998	1	0	0%
1999	5	4	80%
2000	3	1	33%
2001	2	0	0%
2002	6	5	83%
2003	0	0	-
2004	2	0	0%
2005	5	4	80%
2006	6	1	17%
2007	10	5	50%
2008	8	4	50%
	Total: 53	Total: 27	Total: 51%

The recourse to the HUGO gene nomenclature, notably in clinical studies, may increase significantly in the next years as a consequence of the introduction of microarrays and, more generally, bioinformatics in these studies. Indeed, the comparison of microarray data from different sources requires the exact mapping of the names used by different authors. This aim is generally achieved by mapping gene annotations obtained from different microarray platforms to the HUGO gene symbols and names. Many efforts are being made to standardize microarray data. This could ultimately render the HUGO names and symbols more "popular".

As reflected by the present article, the cancer community has not widely adopted the HUGO nomenclature. I suggest that the mention of standard names or symbols in abstracts in place of, or, at least, next to aliases, should be made mandatory. More generally, all actors of research should pay more attention to the usage of standard nomenclature. Indeed, the task of mentioning genes in papers by their official names and symbols will require the collaboration of authors and journals (editors and reviewers). Regarding journals, the pursuit of unique standard gene names and symbols has been encouraged for years by most journals primarily concerned with human genetics or molecular biology (some examples are *American Journal of Human Genetics, Annals of Human Genetics, Biochemical Journal, Bioinformatics, Cancer Genomics and Proteomics, Cytogenetic and Genome Research, Genome Research, Genomics, Human Genetics, Human Mutation, Journal of Medical Genetics, Mammalian Genome, Molecular Therapy, Nature Reviews (all categories), Nucleic Acids Research, Pharmacogenetics and Genomics, PloS Genetics, The Journal of Immunology).* In contrast, the frequency with which the HUGO nomenclature is used is still low in clinical journals that publish most of the data concerning tumor cell identification. The 303 articles examined in the present article were published in more than 100 journals (not listed here), but 177 of them (58%) were published in only 17 journals: *Annals of Oncology, Annals of Surgery, Anticancer Research, Breast Cancer Research and Treatment, Cancer, Cancer Letters, Cancer Research, Clinica Chimica Acta, Clinical Cancer Research, Clinical Chemistry, International Journal of Cancer, International Journal of Oncology, Journal of Cancer Research and Clinical Oncology, Journal of Clinical Oncology, Melanoma Research, Oncology Reports, Surgery*. At the beginning of 2008, none of these journals required or had even proposed the use of HUGO nomenclature to its authors.

Agreement on gene name usage will not only make one's own research easier, but will also be helpful to the present generation (as well as future generations) of researchers who are about to enter molecular clinical research.

Address for correspondence:
Marc Lacroix
InTextoResearch
4, chemin de Hoevel
4837 Baelen, Belgium
e-mail: ITR@iname.com

REFERENCES

1. Pantel K, Brakenhoff RH. Dissecting the metastatic cascade. Nat Rev Cancer 2004; 4: 448-56.
2. Zach O, Lutz D. Tumor cell detection in peripheral blood and bone marrow. Curr Opin Oncol 2006; 18: 48-56.
3. Lacroix M. Significance, detection and markers of disseminated breast cancer cells. Endocr Relat Cancer 2006; 13: 1033-67.
4. White J, Wain H, Bruford E, Povey S. Promoting a standard nomenclature for genes and proteins. Nature 1999; 402: 347.
5. Bruford EA, Lush MJ, Wright MW, Sneddon TP, Povey S, Birney E. The HGNC Database in 2008: a resource for the human genome. Nucleic Acids Res 2008; 36 (Database issue): D445-8.
6. Tamames J, Valencia A. The success (or not) of HUGO nomenclature. Genome Biol 2006; 7: 402.

Received:
Accepted: