

Incorrect use of the term synteny

The term 'synteny' (or syntenic) refers to gene loci on the same chromosome regardless of whether or not they are genetically linked by classic linkage analysis¹. This term was introduced in 1971 by John H. Renwick, of the London School of Hygiene and Tropical Medicine, at the 4th Internal Congress of Human Genetics in Paris with one of us (E.P.) in attendance. The need for such a term was suggested to J.H. Renwick by E.A. Murphy, of Johns Hopkins University². It arose as a consequence of the new methods in gene mapping using somatic cell hybrid cells. Human genes located on the same chromosome with a genetic distance that could not be determined by the frequency of recombination lacked a term of reference.

'Synteny' means 'same thread' (or ribbon), a state of being together in location, as synchrony would be together in time. Although several textbooks³⁻¹⁰ and other reference works¹¹⁻¹⁵ give a correct definition, the term synteny nowadays is often used to refer to gene loci in different

organisms located on a chromosomal region of common evolutionary ancestry. This new usage of the term synteny does not correspond to its original definition and correct language derivation. A survey of 11 articles in *Nature Genetics* since 1992 using the term syntenic or synteny in either the title or the abstract revealed usage incorrect in 8 and ambiguous in 3.

We believe molecular biologists ought to respect the original definition of synteny and its etymological derivation, especially as this term is still needed to refer to genes located on the same chromosome. We recognize the need to refer to gene loci of common ancestry. Correct terms exist: 'paralogous' for genes that arose from a common ancestor gene within one species and 'orthologous' for the same gene in different species.

Eberhard Passarge¹, Bernhard Horsthemke¹ & Rosann A. Farber²

¹*Institut für Humangenetik, Universitätsklinikum Essen, Essen, Germany.*
²*Department of Pathology and Laboratory*

Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. Correspondence should be addressed to E.P. (e-mail: eberhard.passarge@uni-essen.de).

1. Renwick, J.H. *Annu. Rev. Genet.* **5**, 81-120 (1972).
2. Renwick, J.H. in *Human Genetics. Proceedings of the Fourth International Congress of Human Genetics, 6-11 September 1971* (eds de Grouchy, J., Ebling, F.J.G. & Henderson, I.W.) 443-444 (Excerpta Medica, Amsterdam, 1972).
3. Levitan, M. *Textbook of Human Genetics* (Oxford University Press, Oxford, 1988).
4. Sutton, M.E. *An Introduction to Human Genetics* (Harcourt Brace Jovanovich, San Diego, 1988).
5. Thompson, M.W., McInnes, R.R. & Willard, H.F. *Genetics in Medicine* (W.B. Saunders, Philadelphia, 1991).
6. Rothwell, N.V. *Understanding Genetics. A Molecular Approach* (Wiley-Liss, New York, 1993).
7. Connor, J.M. & Ferguson-Smith, M.A. *Essential Medical Genetics* (Blackwell Scientific, Oxford, 1993).
8. Mueller, R.F. & Young, I.D. *Emery's Elements of Medical Genetics* (Churchill Livingstone, Edinburgh, 1995).
9. Jorde, L.B., Carey, J.C. & White, R.L. *Medical Genetics* (Mosby, St. Louis, 1995).
10. Gelehrter, T.D., Collins, F.S. & Ginsburg, D. *Principles of Medical Genetics* (Williams & Wilkins, Baltimore, 1998).
11. Murphy, E.A. & Chase, G.E. *Principles of Genetic Counseling* (Year Book Medical, Chicago, 1975).
12. Rieger, R., Michaelis, A. & Green, M.M. *Glossary of Genetics and Cytogenetics* (Springer Verlag Heidelberg, New York, 1976).
13. Ott, J. *Analysis of Human Genetic Linkage* (Johns Hopkins University Press, Baltimore, 1991).
14. Strachan, T. & Read, A.P. *Molecular Human Genetics* (BIOS Scientific, London, 1996).
15. Snustad, D.P., Simmons, M.J. & Jenkins, J.B. *Principles of Genetics* 749 (John Wiley & Sons, New York, 1997).

Analysis of human transcriptomes

How many human genes are expressed ubiquitously, in all human tissues, and how many are expressed in tissue-specific patterns? To answer these fundamen-

tal questions in molecular biology, we have analysed 3.5-million transcripts from 19 normal and diseased tissue types. We found that as many as 43,500 genes can be

expressed in a single cell type. Only a small fraction of transcripts were exclusively expressed in any individual tissue, whereas nearly 1,000 genes were expressed in all cell types examined. We found 40 genes to be expressed at elevated levels in all cancer tissues but not in normal cells.

Serial analysis of gene expression (SAGE) studies^{1,2} of 84 libraries derived from 19 different sources identified 134,135 transcripts from approximately 84,000 different genes (Table 1; data and analysis available at http://genetics.nature.com/supplementary_info/). Expression levels for these genes ranged from 0.3 to 9,417 transcript copies per cell. The transcript tags matched approximately 4,300 known genes and 41,000 genes with unknown functions, whereas the remaining transcript tags (46%) had no matches to existing databases (Table 2, see http://genetics.nature.com/supplementary_info/).

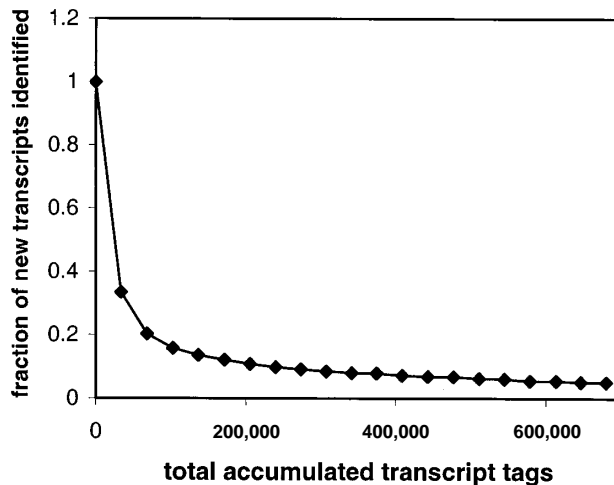
The subset of expression data from colorectal cancer cell lines provided the first relatively complete analysis of the transcripts expressed in a single mammalian cell type. We analysed 643,283 transcripts from colorectal cancer cell lines. As human cells contain approximately 300,000 mRNA molecules, this number was sufficient to provide approximately twofold coverage of the transcriptome, revealing over 83% of transcripts expected to be pre-

Table 1 • Tissues and transcript tags

	Libraries ^b	Total transcripts ^c	Unique genes ^d
Normal tissues^a			
colon epithelium ⁵⁻⁹	2	98,089	12,941
keratinocytes ^e	2	83,835	12,598
breast epithelium ^e	2	107,632	13,429
lung epithelium ¹⁰	2	111,848	11,636
melanocytes ^e	2	110,631	14,824
prostate ^e	2	98,010	9,786
monocytes ^e	3	66,673	9,504
kidney epithelium ^e	2	103,836	15,094
chondrocytes ^e	4	88,875	11,628
cardiomyocytes ^e	4	77,374	9,449
brain ⁹	3	202,448	23,580
Diseased tissues^a			
colon cancer ^{5-9,11,e}	22	1,004,509	56,153
pancreatic cancer ⁵⁻⁸	4	126,414	17,050
breast cancer ^e	5	226,630	18,685
lung cancer ¹⁰	5	221,302	22,783
melanoma ^e	10	269,332	25,600
polycystic kidney disease ^e	2	112,839	16,280
haemangiopericytoma ^e	5	199,985	31,351
brain cancer ⁹	3	186,567	23,108
Total	84	3,496,829	84,103

^aSource of the RNA analysed. ^bThe number of SAGE libraries analysed. ^cThe total number of transcripts analysed from each tissue. ^dThe number of unique genes observed in each tissue. ^eUnpublished data. See Methods (http://genetics.nature.com/supplementary_info/) for the derivation of a number of unique genes.

Fig. 1 Sampling of gene expression in colon cancer cells. Analysis of transcripts at increasing increments of transcript tags indicates that the fraction of new transcripts identified approaches zero at ~650,000 total tags.



sent at levels as low as one copy per cell. We confirmed this prediction by measurement of ascertained new tags at increasing increments of total tags. The fraction of new transcripts emanating from additional SAGE data approached zero at approximately 650,000 tags (Fig. 1).

Expression levels of transcripts in colon cancer cells ranged from 0.5 to 2,672 transcript copies per cell (Table 3, see http://genetics.nature.com/supplementary_info/). The 61 transcripts expressed at over 500 transcript copies per cell made up one-fifth of the mRNA mass of the cell and the most highly expressed 623 genes accounted for nearly one-half of the mRNA content. In contrast, most unique transcripts were expressed at low levels, with just under 23% of the mRNA mass of the cell comprising 90% of the unique transcripts expressed (Table 4, see http://genetics.nature.com/supplementary_info/). A 'virtual rot' analysis of the expressed transcripts, in which the cumulative mRNA content of a cell is plotted as a function of the transcript concentration, identified a relatively continuous distribution of gene expression without discrete abundance classes, similar to those observed in previous rot studies of human cancer cells³ (Fig. 2, see http://genetics.nature.com/supplementary_info/).

Evaluation of SAGE data from other cells revealed that changes in gene expression between physiologic states of a particular cell type or between different samples of the same cell type were less than changes between cell types of different origins (Fig. 3, see http://genetics.nature.com/supplementary_info/). Only a small fraction of transcripts were exclusively expressed in a particular normal or diseased tissue. Detailed analyses of transcripts from epithelia of colon, breast, lung and kidney, melanocytes, and cells from prostate and brain identified transcripts that were nominally

expressed at greater than ten copies per cell in one tissue, but not in any other tissues studied. The fraction of these tissue-specific transcripts ranged from 0.05% in normal prostate to 1.76% in normal colon epithelium (Table 5, see http://genetics.nature.com/supplementary_info/). Approximately 50% of these transcript tags matched known genes or ESTs. Most of these tissue-specific transcripts have not been previously reported in the literature and their roles in the tissues examined provide a wealth of opportunities for further research.

We detected nearly 1,000 transcripts that were expressed at more than or equal to 5 transcript copies per cell in every tissue analysed (Table 6, see http://genetics.nature.com/supplementary_info/). These expressed genes represent a view into the 'minimal transcriptome'—the set of genes expressed in all human cells. Such genes largely represent well-known constitutive or housekeeping genes thought to provide the molecular machinery necessary for basic functions of cellular life⁴. These data may also provide more useful controls for RNA normalization than were formerly available, as the expression profiles of genes previously used for this purpose vary considerably between different tissues and physiologic states.

Finally, we detected 40 genes that were expressed in all cancer tissues examined at levels greater than or equal to 3 transcript copies per cell and whose expression was at least twofold higher in each cancer compared with its corresponding normal tissue (Table 7, see http://genetics.nature.com/supplementary_info/). The observed elevated expression of such genes in many tumour types indicates a potentially general role for these genes in tumorigenesis and suggests they may be useful as diagnostic markers or targets for therapeutic intervention.

The analyses described here provide a heretofore unavailable picture of human transcriptomes. Our results, like the human genome sequence, provide basic information integral to future experimentation of normal and disease biology. As SAGE analyses provide absolute rather than relative expression levels, future SAGE data can be directly integrated with those described here to provide progressively deeper insights into expression patterns.

Acknowledgements

We thank B. Wentworth, D. Jefferson and S.J. de Beur for providing tissues and cell lines; G. Schuler and L. Wagner for their work with the UniGene project; and D. Velculescu for assistance with Monte Carlo simulations. We thank members of our laboratories for useful advice and discussions. This research was supported by NIH grants CA57345, CA62924 and CA43460. B.V. is an investigator of the Howard Hughes Medical Institute. K.W.K. received research funding from Genzyme Molecular Oncology.

Victor E. Velculescu¹, Stephen L. Madden⁵, Lin Zhang¹, Alex E. Lash⁶, Jian Yu¹, Carlo Rago², Anita Lal⁷, Clarence J. Wang⁵, Gary A. Beaudry⁵, Kristin M. Ciriello⁵, Brian P. Cook⁵, Michael R. Dufault⁵, Anne T. Ferguson¹, Yuhong Gao⁵, Tong-Chuan He¹, Heiko Hermeking², Siewleng K. Hiraldo⁵, Paul M. Hwang^{1,2,3}, Marissa A. Lopez⁵, Hilary F. Luderer⁵, Brynna Mathews⁵, Joseph M. Petroziello⁵, Kornelia Polyak⁸, Leigh Zawel^{1,2}, Wen Zhang⁵, Xiaoming Zhang⁵, Wei Zhou^{1,2}, Frank G. Haluska⁹, Jin Jen⁴, Saraswati Sukumar¹, Gregory M. Landes⁵, Gregory J. Riggins⁷, Bert Vogelstein^{1,2} & Kenneth W. Kinzler¹

¹The Johns Hopkins Oncology Center and ²The Howard Hughes Medical Institute, ³Division of Cardiology, ⁴Division of Otolaryngology, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. ⁵Genzyme Molecular Oncology, Framingham, Massachusetts, USA. ⁶National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, Maryland, USA. ⁷Department of Pathology, Duke University School of Medicine, Durham, North Carolina, USA. ⁸Department of Adult Oncology, Dana-Farber Cancer Institute, and ⁹Division of Hematology/Oncology, Massachusetts General Hospital, Boston, Massachusetts, USA.

Correspondence should be addressed to V.E.V. (e-mail: velculescu@jhmi.edu).

1. Velculescu, V.E., Zhang, L., Vogelstein, B. & Kinzler, K.W. *Science* **270**, 484–487 (1995).
2. Velculescu, V.E. et al. *Cell* **88**, 243–251 (1997).
3. Bishop, J.O., Morton, J.G., Rosbash, M. & Richardson, M. *Nature* **250**, 199–204 (1974).
4. Lewin, B. *Gene Expr.* **2**, 694–727 (1980).
5. Polyak, K., Xia, Y., Zweier, J.L., Kinzler, K.W. & Vogelstein, B. *Nature* **389**, 300–304 (1997).
6. Zhang, L. et al. *Science* **276**, 1268–1272 (1997).
7. Hermeking, H. et al. *Mol. Cell* **1**, 3–11 (1997).
8. He, T.C. et al. *Science* **281**, 1509–1512 (1998).
9. Lal, A. et al. *Cancer Res.* **59**, 5403–5407 (1999).
10. Hibi, K. et al. *Cancer Res.* **58**, 5690–5694 (1998).
11. Yu, J. et al. *Proc. Natl Acad. Sci. USA* (in press).