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 analysis. This term was introduced in 1971 by John H. Renwick, of the London School of Hygiene and Tropical Medicine, at the 4th International 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 University. 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.

'Synten' means 'same thread' (or ribbon), a state of being together in location, as synchrony would be together in time. A state of being together in location, bon), a state of being together in location, synchrony would be together in time. Although several textbooks and other reference works give a correct definition, the term synteny nowadays is often used to refer to gene loci in different species. The term 'paralogous' for genes that arose from a common ancestor gene within one species and 'orthologous' for the same gene in different species. We recognize the need to refer to gene loci in different species. 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
¹Institut für Humangenetik, Universitätsklinikum Essen, Essen, Germany.
²Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

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 fundamental 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 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 transcriptome 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-
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 ascertainment 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 cells (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 opportunites 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 life. 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.I. de Bear 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. Madden3, Lin Zhang1, Alex E. Lash3, Jian Yu1, Carlo Rago1, Anita Lal1, Clarence J. Wang5, Gary A. Beaudry1, Kristin M. Ciriello5, Brian P. Cook3, Michael R. Dufault3, Anne T. Ferguson1, Yuhong Gao3, Ting-Chuan He1, Heiko Hermeking2, Siewleng K. Hiraldo2, Paul M. Hwang1,2,3, Marissa A. Lopez5, Hilary F. Luderer5, Brynna Mathews1, Joseph M. Petroziello5, Kornelia Polyak3, Leigh Zawel1,2,3, Wen Zhang5, Xiaoming Zhang5, Wei Zhou1,2,3, Frank G. Haluska2, Jin Jen4, Saraswati Sukumar1, Gregory M. Landes5, Gregory J. Riggins1,2, Bert Vogelstein1,2 & Kenneth W. Kinzler1

1The Johns Hopkins Oncology Center and 2Division of Cardiology, 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. (e-mail: velculescu@jhmi.edu).

Synteny mapping with T. aestivum monosomics (or monotelosomics) exploits the altered segregation ratio that characterizes progeny of a monosomic compared to a disomic. In the monosomic portion of progeny from a cross between a monosomic female and euploid male, the monosome is contributed by the male parent. To answer this question, we turn to the dotplot capabilities of the Synteny Database. This implementation starts with the first gene on an index chromosome, here Hsa17, and then looks for that gene’s zebrafish ortholog or co-orthologs. When it finds an ortholog, it puts a point on the zebrafish chromosome directly vertical to the human gene.