

USAGE: a web-based approach towards the analysis of SAGE data

A. H. C. van Kampen^{1,*}, B. D. C. van Schaik¹, E. Pauws², E. M. C. Michiels³, J. M. Ruijter⁴, H. N. Caron⁵, R. Versteeg⁵, S. H. Heisterkamp⁶, J. A. M. Leunissen⁸, F. Baas⁷ and M. van der Mee¹

¹Bioinformatics Laboratory, ²Laboratory of Pediatric Endocrinology, ³Department of Pediatric Oncology, ⁴Department of Anatomy and Embryology, ⁵Department of Human Genetics, ⁶Department of Clinical Epidemiology and Biostatistics, ⁷Neurozintuigen Laboratory, Academic Medical Center, Meibergdreef 9, 1100 AZ Amsterdam, The Netherlands and ⁸Centre for Molecular and Biomolecular Informatics, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

Received on April 11, 2000; accepted on May 22, 2000

Abstract

Motivation: SAGE enables the determination of genome-wide mRNA expression profiles. A comprehensive analysis of SAGE data requires software, which integrates (statistical) data analysis methods with a database system. Furthermore, to facilitate data sharing between users, the application should reside on a central server and be accessed via the internet. Since such an application was not available we developed the USAGE package.

Results: USAGE is a web-based application that comprises an integrated set of tools, which offers many functions for analysing and comparing SAGE data. Additionally, USAGE includes a statistical method for the planning of new SAGE experiments. USAGE is available in a multi-user environment giving users the option of sharing data. USAGE is interfaced to a relational database to store data and analysis results. The USAGE query editor allows the composition of queries for searching this database. Several database functions have been included which enable the selection and combination of data. USAGE provides the biologist increased functionality and flexibility for analysing SAGE data.

Availability: USAGE is freely accessible for academic institutions at <http://www.cmbi.kun.nl/usage/>. The source code of USAGE is freely available for academic institutions on request from the first author.

Contact: a.h.vankampen@amc.uva.nl

Introduction

Serial Analysis of Gene Expression (SAGE; Velculescu *et al.*, 1995) is a technique to quantitatively analyse

genome-wide gene expression profiles (Madden *et al.*, 1997; Zhang *et al.*, 1997; Michiels *et al.*, 1999; Velculescu *et al.*, 1999). SAGE is based on the principle that a short sequence tag (9–10 bp) derived from a specified position, e.g. the most 3' NlaIII site, in a mRNA sequence is in principle sufficient to uniquely identify a transcript. A SAGE experiment results in the generation of a large set of long serial DNA molecules (concatemers) consisting of ditags, which are punctuated by the four base CATG sequence. The frequency, with which a particular tag is observed in these concatemers, provides the relative expression level of the corresponding transcript. The sequenced concatemers are the starting point for data analysis. The first step of the analysis is the extraction of ditags from the concatemer sequences. From these ditags a list of tags is compiled, which we will call the SAGE tag list. The next step in the analysis is the determination of the transcripts from which the tags were derived in order to identify the genes that were expressed. Finally, transcript frequencies in two or more SAGE tag lists can be (statistically) compared to distinguish differences in gene expression in the respective tissues or cells.

Table 1 summarises the public resources that are available for the analysis of SAGE data. The SAGE300 program (Zhang *et al.*, 1997) is probably the most commonly used application for SAGE analysis. In order to identify SAGE tags this program compiles a database of tags extracted from human (EST) sequences in Genbank. A drawback of this method is that the orientation of the sequence is not checked before tag extraction, consequently, incorrect tags can result. As part of the Cancer Genome Anatomy Project (CGAP; Strausberg

*To whom correspondence should be addressed.

et al., 1997), in which SAGE is used to identify genes that undergo alteration in expression during malignant transformation, NCBI established the SAGEmap public database (Lal *et al.*, 1999). In SAGEmap, a tag-to-gene mapping is built on a regular basis and is an improvement over the approach taken in the SAGE300 program, because more efforts are taken to define the orientation of the sequence prior to tag extraction. In addition to this standard tag-to-gene mapping, a 'reliable tag-to-gene mapping' is also constructed which accounts for sequencing errors in Genbank sequences. These two tag-to-gene mappings can be downloaded and used in combination with applications such as Microsoft Access. Alternatively, the tag-to-gene mappings are accessible online from the SAGEmap site but this allows only the analysis of one tag at a time. No full identification reports, i.e. for all tags in a SAGE tag list, can be generated as is possible with SAGE300, which unfortunately does not support the use of these tag-to-gene mappings. The SAGEmap site, however, offers several additional utilities such as the 'Virtual Northern' to extract SAGE tags and orientation signals from a sequence, the 'xProfiler' for pooling and comparison of SAGE tag lists present in SAGEmap and a gene-to-tag mapping database. A third SAGE resource comprises the online query tool of the *Saccharomyces Genome Database* (SGD), which allows users to search and present SAGE data obtained for yeast (Velculescu *et al.*, 1997) in several manners.

The (statistical) comparison of genome-wide expression data is still a subject of debate (Claverie, 1999; Vingron and Hoheisel, 1999), and this explains why several statistical approaches were developed for the comparison of SAGE tag lists (Madden *et al.*, 1997; Audic and Claverie, 1997; Zhang *et al.*, 1997; Chen *et al.*, 1998; Kal *et al.*, 1999; Michiels *et al.*, 1999). We recently compared several methods, which showed that the available tests essentially gave similar results (publication in preparation). Most of these statistical tests are publicly available in the form of a software application, or are part of a SAGE analysis software package (see Table 1).

Although the abovementioned tools and databases for the analysis of SAGE data are freely available, none of these resources individually offers the functionalities required by a user to perform a complete and comprehensive SAGE analysis. Therefore, in practice it is necessary to combine tools and databases from different resources. In this paper we introduce the USAGE application, which offers a full set of methods for the extraction of tags from concatemers, the identification of tags, the (statistical) comparison of SAGE tag lists, and presentation of results. All these methods are integrated in one web-enabled and multi-user environment. USAGE borrows ideas from SAGE300, uses the tag-to-gene mappings from NCBI and public SAGE tag lists from CGAP, and implements

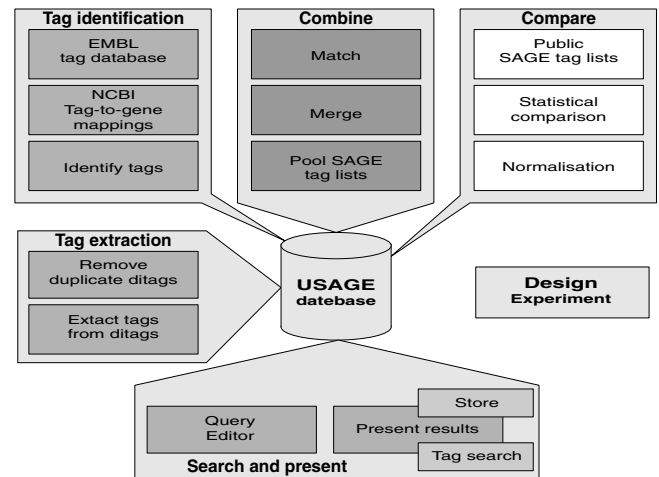


Fig. 1. Schematic overview of databases and methods that are integrated in the USAGE program. The methods for the design of future experiments do not make use of the USAGE database.

the statistical method described in Kal *et al.* (1999). In addition, as described below, several other functionalities are offered by USAGE that are not found in the other programs. The USAGE application is completely transparent and self-contained and its automated data conversion facilities permit the biologist to focus on data analysis and interpretation of data.

System and methods

The USAGE program is a web-application that runs on a Unix server and is accessed after providing a username and password. This ensures that a user can only access and use his/her own data and data denoted as public within the USAGE environment. Before starting data analysis, USAGE uploads the concatemers, which are to be analysed, to the Unix server. All concatemers that correspond to the same SAGE library must be stored in separate text files, which subsequently should be added to a single archive file (zip, gzip or tar). The most important functionalities of USAGE are shown in Figure 1 and discussed below.

The USAGE database

The USAGE database is a relational database that comprises several tables which contain the tags and tag counts, duplicate ditags, tag-to-gene mappings, tag identifications, results from statistical comparisons (probability values), results from previous queries, and results from the pool, match and merge options (see below). As part of the USAGE database, a table catalog is maintained that contains meta information about the available tables (e.g. SAGE tag lists or tag identifications) and their attributes (e.g. tag, tag

Table 1. Public resources available for the analysis of SAGE data

Resource	Main functionalities	Web site
SAGE300 (Zhang <i>et al.</i> , 1997)	Tag extraction, tag identification, statistical comparison	www.sagenet.org/
NCBI SAGEmap (Lal <i>et al.</i> , 1999)	Tag identification, statistical comparisons of CGAP SAGE tag lists (xProfiler), Virtual Northern	www.ncbi.nlm.nih.gov/SAGE/
Significance test (Audic and Claverie, 1997)	Statistical comparison of pairs of tags	igs-server.cnrs-mrs.fr/
Significance test (Kal <i>et al.</i> , 1999)	Statistical comparison of pairs of tags and planning of new SAGE experiments	Email: j.m.ruijter@amc.uva.nl
Saccharomyces Genome Database (Velculescu <i>et al.</i> , 1997)	Online queries of yeast SAGE data, and yeast tag database	genome-www.stanford.edu/ cgi-bin/SGD/SAGE/querySAGE

count, accession code). The table catalog includes information such as the project name, data owner, tag length and anchoring enzyme used during tag extraction. If an attribute (tag, accession code and Unigene cluster identifier) of a table can be linked to an external web-site the table catalog includes information (the syntax of the URL) to establish this connection.

Extraction of ditags and tags from the concatemers

Prior to ditag and tag extraction from the concatemer sequences the anchoring enzyme (e.g. NlaIII), the tag length (e.g. 10) and the minimum and maximum ditag lengths (e.g. 20 and 24, respectively) are set by the user. Ditags that are too short or too long are discarded from further analysis, as these are considered experimental artefacts. Duplicate ditags are removed for the same reason, since the event of a duplicate ditag occurrence by chance is generally very low. Subsequently, a SAGE tag list (tag and tag counts) is compiled from the remaining ditags. The (duplicate) ditags, tags and tag counts are stored in the USAGE database and are complemented with information about the concatemer(s) from which they were extracted, which allows backtracking of the tag extraction process.

The tag databases and tag identification

USAGE includes NCBI's 'tag-to-gene mapping' and 'reliable tag-to-gene mapping' for tag identification. In addition, it includes the yeast tag database that is part of the Saccharomyces Genome Database. Alternatively, the user may generate a tag database from the EMBL nucleotide database. Since the EMBL database is subdivided into several sections (e.g. human, rodent, EST, etc.), the section from which the tag database is to be compiled is selected first. Subsequently, from each EMBL database sequence in the selected section, the tag located 3'-adjacent to the most 3' NlaIII site, is extracted. This tag is complemented with the sequence accession code,

description, and six base trailer following the tag. The six base trailer helps to discriminate identical tags that were extracted from different sequences. USAGE also offers the possibility to compile a tag database for one or more organisms (e.g. *H. sapiens* and *S. cerevisiae*) in which case all EMBL sequences, with exception of the sequences in the EST, STS, GSS and HTS sections, are considered for tag extraction. In contrast to the procedure followed to generate the NCBI tag-to-gene mapping, no efforts are made to check sequence orientation. However, an option is available that allows tag extraction from a sequence in both the normal (i.e. as included in the database) and complement-reverse orientation. Once a tag database is compiled it is available to every user of USAGE. Consequently, when tags need to be identified, the user may decide to compile a new tag database or use one of the existing databases. SAGE tags are identified by matching them against the tags in the tag database. The result of this identification process is automatically stored in the USAGE database. The method that is implemented for extracting tags from the EMBL database sequences is basically identical to the method implemented in the SAGE300 program with the exception that USAGE allows the extraction of tags from the complement-reverse orientation of the database sequence.

Public SAGE tag lists

The comparison of SAGE tag lists may provide important information about transcripts that are specific for a cell, tissue or (pathological) condition. Therefore, USAGE supports the use of the CGAP public SAGE tag lists that were obtained from various tissues. Furthermore, other tag lists can also be denoted as public on request of the user. These tag lists then become automatically available to every user for comparison with other lists.

Normalisation of SAGE tag lists

Normalisation of tag counts facilitates the comparison of differently sized SAGE tag lists via visual inspection. Normalisation comprises the equalisation of the total number of tags in each list by adjusting the tag counts in the individual lists. Each SAGE tag list can be normalised to one of four predefined list sizes (10 000, 50 000, 100 000 and 1 000.000). During normalisation all tag counts are adjusted by multiplying each tag count with L/N , where L is the predefined size and N is the size of the SAGE tag list that is normalised. The final normalised tag counts are rounded to the nearest integer. It is important to note that normalisation should never precede the statistical comparison of SAGE tag lists because adjustment of list sizes artificially affects the detectable difference. Comparison of tag counts after normalisation only gives a first impression of differences between SAGE tag lists and does not predict anything about statistical significance.

Statistical comparison of SAGE tag lists

USAGE implements the statistical test described by Kal *et al.* (1999), which is based on the assumption that the tag counts are binomially distributed, and can be used to compare SAGE tag lists of unequal size. To start the comparison, two or more SAGE tag lists are selected from which all pairwise combinations of lists are generated. Subsequently, all tags in each pair of lists are statistically compared and the resulting probability values are stored in the USAGE database. A 'probability value' of -2 results if a particular tag is not present in either of the paired lists but is included in at least one of the selected SAGE tag lists. In this way, during presentation of the results, one can easily determine the pairwise combinations in which a particular tag was not included. The statistical results can now be queried from the query editor (see below) to identify the most interesting tags for further investigation.

Planning of future SAGE experiments

The significance test described in Kal *et al.* (1999) is easily inverted and used for the planning of new SAGE experiments. USAGE allows the calculation of several quantities:

- (1) Given the sizes of two SAGE tag lists USAGE calculates the detectable difference between tag counts for a chosen significance level and power of the test.
- (2) Given an observed difference in tag counts and the total number of tags sequenced in both SAGE tag lists and the chosen significance level, USAGE calculates the power of the test can be determined.
- (3) Given an expected difference in tag counts and a chosen significance level, a required power, USAGE

Fig. 2. The query editor provides a flexible way to select and search SAGE data. The form is automatically generated and is based on the information contained in the selected table.

date	trailer	idline	class	accession
				srs
19980615	gacgac	Homo sapiens 22 kDa actin-binding protein (SM22) gene, complete cds	Class C (sense)	AF013711
19970827	gacgac	Homo sapiens DNA for SM22 alpha, complete cds	Class C (sense)	D84342
19970821	gacgac	Homo sapiens mRNA for SM22 alpha, complete cds	Class A (sense)	D17409
19950106	gacgac	Human SM22 mRNA, 5' end	Class A (sense)	M83106
19950106	gacgac	Human 22kDa smooth muscle protein (SM22) mRNA, complete cds	Class A (sense)	M95787

Fig. 3. The results of the identification of tag 'ACAGGCTACG'. The accession codes are linked to the EMBL nucleotide database.

calculates the minimum number of tags needed in each list such that for this difference the statistical significance can be reached.

- (4) Given an expected difference in tag counts and a significance level and a power and the number of tags already sequenced in an existing SAGE tag list, USAGE calculates the number of tags needed in a new list such that for this difference the statistical significance can be reached.

Table 2. Two hypothetical SAGE tag lists (a) are used to demonstrate the match and merge functions (N denotes the tag counts). Table 2b presents the result of the 'match' operation applied to list 2 against list 1. The result is a list of all tags and counts of the first list and the corresponding tag and tag count from the second list. Note that matching list 1 against 2 is not identical to this match of list 2 against 1. Table 2c presents the result of the 'merge' operation applied to list 1 and 2

Table 2a

List 1		List 2	
Tag 1	N1	Tag 2	N2
AATACCCCA	1	AATACCCCA	5
ATTGTTGAAT	10	GGCCGAAAAT	2
TGGGCGCGCT	9	ATTGTTGAAT	11
CCGTACCAAT	4	CCGTACCAAT	2
Total	24		20

Table 2b

Match lists			
Tag 1	N1	Tag 2	N2
AATACCCCA	1	AATACCCCA	5
ATTGTTGAAT	10	ATTGTTGAAT	11
TGGGCGCGCT	9		
CCGTACCAAT	4	CCGTACCAAT	2
Total	24		18

Table 2c

Merged lists			
Tag 1 + 2	N1		N2
AATACCCCA	1		5
ATTGTTGAAT	10		11
TGGGCGCGCT	9		0
CCGTACCAAT	4		2
GGCCGAAAAT	0		2
Total	24		20

The query editor

The query editor allows the user to search and present his data and results from analysis in a very flexible manner. From a technical point of view, the query editor is a web-interface that allows the user to compose SQL queries for the USAGE database. Once a table is chosen from the USAGE main menu and the query editor (Figure 2) is entered, a web form is automatically generated from information contained in the table catalog. Each pull-down box on this form contains the names of all attributes of the selected table, which allows the user to compose a

presentation layout. Several functions are included and allow the calculation and presentation of sums, differences or (log) ratios of tag counts for two or more selected SAGE tag lists. To avoid each query resulting in a dump of all information to a table, the user first composes a summary presentation for the (unique) tags in a selected table, through the selection of a limited number of attributes. Subsequently, a set of attributes is selected that are presented for individual tags in a 'full view' layout (Figure 3). If the summary presentation is disabled then for all tags the selected attributes are presented immediately in the full view. To enable the selection of a subset of the data, a range of operators (greater than, equal to, matching, etc.) is available that can be applied to any attribute in the table. The results of a query can be sorted on any attribute in descending or ascending order, and also the maximum number of records to show in the full view can be limited. The results of a query are displayed in the browser and can be printed or exported as a tab-delimited file for further processing with other programs.

Storage of query results

An important feature of USAGE is that all results of a particular query can be stored in the USAGE database. In this way sub-selections (e.g. only tags that occur more than ten times) or results from normalisation can be saved for further processing. USAGE automatically creates a new table when a user stores query results. In addition, the table catalog is updated with (user defined) information about the origin of the data in the new table. This ensures that the new table is available for queries or to be used with the pool, match or merge option.

Tag search

If one suspects that a particular tag is the result of a sequencing error, the tag search option of USAGE offers the possibility to retrieve similar tags in the list. Tags are considered similar if they have substituted bases at only a limited number of positions (typically one or two). The query tag and the maximum number of allowed substitutions are the input for this tag retrieval function. The tag search function does not take base insertions or deletions into account.

Pool SAGE tag lists

When multiple SAGE tag lists are available from the same tissue it is sometimes useful to pool these lists into one 'new' list in which the tag counts for identical tags are added. For example, one could pool SAGE tag lists that were obtained from brain tumor tissue in one group and pool SAGE tag lists that were obtained from normal brain tissue in a second group. The resulting two lists are automatically added to the USAGE database, and can be subjected to further analysis. These pooled lists

Table 3. Components required by USAGE

Component	Internet site
USAGE core program	Available from author on request
EMBL database	ftp://ftp.ebi.ac.uk/pub/databases/embl/release/ ftp://ftp.ebi.ac.uk/pub/databases/embl/new/
NCBI tag-to-gene mappings	ftp://ncbi.nlm.nih.gov/pub/plash/
CGAP public SAGE tag lists	www.ncbi.nlm.nih.gov/SAGE/sagelb.cgi
Perl and Perl modules	www.perl.org
Postgres DBMS	www.postgresql.org
R	www.ci.tuwien.ac.at/R/contents.html
Apache web-server	www.apache.org
GNU sort	ftp://ftp.gnu.org/gnu/textutils/
Yeast SAGE database	ftp://genome-ftp.stanford.edu/pub/yeast/tables/SAGE_Data/

may be merged with the original lists in order to include information about the contributions of the original lists to the pooled tag count.

Match and merge of data

The match and merge functions allow the combination of any USAGE database table that have an attribute (e.g. tag) in common. The match function is, for example, used during tag identification in which case the tags from the SAGE tag list are matched against a tag database. In contrast, the merge function fully combines two tables and can, for example, be used to combine normalised SAGE tag lists with results from statistical analysis. Table 2 demonstrates the use of the match and merge functions for two hypothetical SAGE tag lists. The result of a match or merge operation is automatically stored in the USAGE database and the table catalog is updated. This new table is then available for further analysis. In practice it turns out that the combination of match and merge is required to make appropriate combinations of tables stored in the USAGE database.

Implementation

The USAGE program is a web application with a client/server architecture that runs on a Unix server and is accessed via a web browser (e.g. Netscape) from any computer that has access to the server via the internet or local intranet. USAGE was developed on a SUN platform (Solaris 2.5) and is predominantly written in Perl. USAGE consists of a core Perl program that implements most of its functionalities, but which also interfaces all other components. The statistical procedures of USAGE were implemented in the 'R' statistical package. All components that are required to setup USAGE are freely available (Table 3). The GNU sort application is required in order to sort large data files. The Perl CGI and DBI modules are used to interact with the web-browser and the Postgres DBMS respectively. The Apache web

server is used to make USAGE accessible via a web browser.

Discussion

With USAGE we developed a large-scale web-application that enables a comprehensive analysis of SAGE data with regard to tag identification and (statistical) comparison of SAGE tag lists. The combination of the query editor and the match, merge and pool functionalities offers a powerful and flexible approach towards data management and consequently, allows the biologist to present and analyse his or her data in the most appropriate way. Nevertheless, USAGE can be improved and extended in several ways. At this moment USAGE lacks adequate methods for dealing with sequencing errors, which can be present in SAGE tags or tags extracted from database sequences. Sequencing errors can seriously complicate the analysis of the data. In a next version of the application we will include functionalities such as described in Velculescu *et al.* (1999) for identifying sequencing errors in SAGE tags. Furthermore, in order to deal with sequencing errors in tags extracted from database sequences, we are developing an improved tag-to-gene mapping. Incorrect tags may also result from sequencing errors that destroy the most 3' CATG or introduce a false CATG. Furthermore, problems in the recognition of the 3' end of EST database sequences may yield tags from the wrong cDNA strand. Our improved tag-to-gene mapping will also account for these error sources. A paper describing this mapping is in preparation. Another extension of USAGE comprises the incorporation of a list of confirmed SAGE tags, which are tags for which the corresponding gene is known and verified. Such a list will further enhance the tag identification process. Finally, we are developing the program BATS (Bayesian Analysis of Tag Series), which does not presume a binomial distribution of SAGE tags, for the statistical comparison of SAGE tag lists. Furthermore, BATS includes a method for clustering SAGE data to identify tags with similar expression

profiles. BATS will be integrated with USAGE in a next version. However, a first version of BATS is already available from the author as a Splus program.

Acknowledgements

USAGE includes the SAGEmap tag-to-gene mapping databases that were developed at the NCBI (Bethesda, USA). USAGE includes SAGE tag lists that became available from the CGAP project. USAGE borrows ideas from the SAGE300 program. We thank A.C.M. Luyf for checking and comparing several tag databases. R.S. Wolfram, E.H. Beekman and J. Haak (department for Information and Communication Technology AMC, Amsterdam) are acknowledged for their assistance concerning several technical implementation issues. We thank N.E. Hain for several improvements of this text.

References

- Audic,S. and Claverie,J.M. (1997) The significance of digital gene expression profiles. *Genome Res.*, **7**, 986–995.
- Chen,H., Centola,M., Altschul,S.F. and Metzger,H. (1998) Characterization of gene expression in resting and activated mast cells. *J. Exp. Med.*, **188**, 1657–1668.
- Claverie,J.M. (1999) Computational methods for the identification of differential and coordinated gene expression. *Hum. Mol. Gen.*, **8**, 1821–1832.
- Kal,A.J., van Zonneveld,A.J., Benes,V., van de Berg,M., Groot Kerkamp,M., Albermann,K., Strack,N., Ruijter,J.M., Richter,A., Dujon,B., Ansorge,W. and Tabak,H.F. (1999) Dynamics of gene expression revealed by comparison of serial analysis of gene expression transcript profiles from yeast grown on two different carbon sources. *Mol. Biol. Cell*, **10**, 1859–1872.
- Lal,A., Lash,A.E., Altschul,S.F., Velculescu,V., Zhang,L., McLendon,R.E., Marra,M.A., Prange,C., Morin,P.J., Polyak,K., Papadopoulos,N., Vogelstein,B., Kinzler,K.W., Strausberg,R.L. and Riggins,G.J. (1999) A public database for gene expression in human cancers. *Cancer Res.*, **59**, 5403–5407.
- Madden,S.L., Galella,E.A., Zhu,J., Bertelsen,A.H. and Beaudry,G.A. (1997) SAGE transcript profiles for p53-dependent growth regulation. *Oncogene*, **15**, 1079–1085.
- Michiels,E.M.C., Oussoren,E., van Groenigen,M., Pauws,E., Bossuyt,P.M.M., Voute,P.A. and Baas,F. (1999) Genes differentially expressed in medulloblastoma and fetal brain. *Physiol. Genomics*, **1**, 83–91.
- Strausberg,R.L., Dahl,C.A. and Klausner,R.D. (1997) New opportunities for uncovering the molecular basis of cancer. *Nat. Genet.*, **15**, 415–416.
- Vingron,M. and Hoheisel,J. (1999) Computational aspects of expression data. *J. Mol. Med.*, **77**, 3–7.
- Velculescu,V.E., Zhang,L., Vogelstein,B. and Kinzler,K.W. (1995) Serial Analysis of Gene Expression. *Science*, **270**, 484–487.
- Velculescu,V.E., Zhang,L., Zhou,W., Vogelstein,J., Basrai,M.A., Bassett,D.E., Hieter,P., Vogelstein,B. and Kinzler,K.W. (1997) Characterization of the yeast transcriptome. *Cell*, **88**, 243–251.
- Velculescu,V.E., Madden,S.L., Zhang,L., Lash,A.E., Yu,J., Rago,C., Lal,A., Wang,C.J., Beaudry,G.A., Ciriello,K.M., Cook,B.P., Dufault,M.R., Ferguson,A.T., Gao,Y., He,T.C., Hermeking,H., Hiraldo,S.K., Hwang,P.M., Lopez,M.A., Luderer,H.F., Mathews,B., Petroziello,J.M., Polyak,K., Zavel,L., Zhang,W., Zhang,X., Zhou,W., Haluska,F.G., Jen,J., Sukumar,S., Landes,G.M., Riggins,G.J., Vogelstein,B. and Kinzler,K.W. (1999) Analysis of human transcriptomes. *Nat. Genet.*, **23**, 387–388.
- Zhang,L., Zhou,W., Velculescu,V.E., Kern,S.E., Hruben,R.H., Hamilton,S.R., Vogelstein,B. and Kinzler,K.W. (1997) Gene expression profiles in normal and cancer cells. *Science*, **276**, 1268–1272.