

FUNCTIONAL ANALYSIS OF GENES

Dorota SITNICKA, Katarzyna FIGURSKA, Sławomir ORZECZOWSKI

Department of Biochemistry, Faculty of Agriculture and Biology, Warsaw
University of Life Sciences-SGGW

DOI: 10.2478/v10052-010-0001-y

Summary: The aim of this article is to present the current literature concerning the expression analysis and methods of functional characteristics of genes. The progress in the analysis of gene expression within cells or whole tissues is undisputed and leads to a constant improvement of our understanding of the function of particular gene. The traditional methods of the functional characteristics of genes such as homology, inactivation and overexpression are more frequently being replaced by microarray and DNA chip analysis, which are extensively supported by bioinformatics tools. Knowledge of the functions and changes in gene expression has applications in medical diagnostics, the pharmaceutical industry and in plant and animal biotechnology.

Key words: gene expression, gene, DNA chips, microarray

INTRODUCTION

The spectacular reports of completely sequenced genomes of yet another organism have been there since 2000, when the weekly "Nature" magazine published the full sequence of the model plant organism - *Arabidopsis thaliana* [1]. Current literature shows that on average, once a month, global science is enriched with full sequences of four genomes of new organisms [5]. However, the recognition of the DNA sequence is merely the beginning on the way to understanding how an organism functions at the molecular level [7]. The identification of coding sequences within a genome and an analysis of gene products are the next phases leading to the recognition of the functions fulfilled by genes. The analysis of gene function, is a task for the newly emerging field functional genomics. There are numerous methods of characterising gene functions

and their number is constantly increasing due to quicker and quicker progress of technology and computerisation.

The process of determining the function of a studied gene is only possible after obtaining the sequence and identifying where it is expressed [7]. It is performed by means of computer analyses or experimental studies, associating a gene and related phenotype. Various strategies and research techniques, such as homology-based prediction, increased expression or gene inactivation have been used for many years now. However, more and more modern microarrays and DNA chips are used and they have become the basic technique of analysing gene function [27].

METHOD OF PREDICTING GENE FUNCTION THROUGH HOMOLOGY

Communal access to genome sequences, which are stored in databases, allows for a comparison with the studied sequences. An analysis showing similarity with the recognised genes is helpful in the initial determination of a probable gene product and its function [12].

The method of predicting through homology relies on the assumption that if a newly sequenced gene is very similar to an already characterized and published gene, the function of the new gene is probably similar.

In this method the first step is comparing the studied fragment of DNA with the fragments of DNA of model organisms available in the computer databases. DNA of most model organisms have been completely sequenced and at least partially characterised with regard to gene function. The largest and most popular databases, in which full sequences of genomes have been stored, are on the NCBI (National Center for Biotechnology Information) server. The biggest limitation in the use of homology-based methods is the presence of uncharacterised sequences.

A comparison between an analysed sequence and described genes contained in the databases also has certain risks. Although at the nucleotide level the DNA sequences may have high similarity, they can encode different amino acids, as well as proteins of various functions and structures. A protein is comprised of many amino acids arranged in an specific order, thus a rule has been adopted that the comparison uses not the sequences of nucleotides, but the amino acids coded by them. It is known that the amino acids sequence is more conserved than the nucleotide one. Use of the amino acids sequence allows detecting the shift of reading frame between the non-homologous genes. Thus, the fragments for comparison are selected in the databases on the basis of the highest possible similarity of the content and order of amino acids.

Using a technique based on homology we can also determine the evolutionary relations between the studied and model organisms [12]. The method of predicting functions by means of homology is the basis of phylogenetics, which compares the sequences based on the common, evolutionary origin of genes. Created phylogenetic trees simplify the search for gene functions, considering their duplications, substitutions in the sequence of various species leading to divergence, and resulting from that - speciation. The knowledge concerning the way of gene flow and determining related species contributes to the initial identification of gene function. On the basis of similarity of gene sequences we can conclude that homologous genes, that is, ones having a common evolutionary ancestor and occurring in the studied organisms, probably have the same or comparable function. Unfortunately, a difficulty lies in the fact that with time, genes may accumulate random mutations and though being highly similar in structure, they may have distinct and separate functions [39]. Based on that we can distinguish between:

- **Orthologs**, are genes occurring in various species (which may also prove that they have occurred in a common ancestor) and fulfilling the same or comparable functions. The genes are created as a result of gene duplication leading to speciation, but do not necessarily have the same function as the gene of a common ancestor [17];
- **Paralogs**, are genes occurring in various organisms or just in one, but due to slight, significant changes in structure, they fulfill separate roles. In this case the duplication leads to divergence, that is, a division of functions. An example of two paralogs are human myoglobin and hemoglobin, responsible for the storage of oxygen in skeletal muscles and transport of oxygen between cells and pulmonary alveoli respectively [17];
- **Xenologs**, are genes similar to one another due to the fact that they have been acquired by organisms through horizontal gene transfer, which does not, however, prove their common evolutionary origin [36].

The requirement for the knowledge concerning gene functions and the development of bioinformatic technologies has caused many scientific institutions to engage into cataloguing knowledge of known homologues in databases (Table 1) [12]. This has allowed for the grouping of data and common access to it.

Researchers comparing the genomes of alga *Chlamydomonas* with the genome of humans and model plant species- *Arabidopsis thaliana* - used the homology-based method of determining the gene function and BLAST algorithm. The aim of the experiment was to determine the relationship between plant and animal kingdoms and to assign the characteristic plant and animal gene functions. As a result, 349 plant proteins engaged in the process of photosynthesis and 195 animal proteins responsible for the movement described [31].

The use of the BLAST algorithm while studying homology is limited when the similarity between the studied sequences is low (reaching 20-30%) [19]. Such proteins that differ significantly at the amino acid level can however, assume a similar structure, fulfill similar functions and also be homologues. In common classification, where proteins have similar sequence, structure and function, they are combined into families, and then into super-families. These proteins that differ in terms of sequences are described as so-called distant homologues. The studies on evolution show that the structure of proteins is preserved better than their sequence [16]. Due to this, studies on protein structure are important when determining their functions. Accordingly, the requirement for the creation of methods studying distant homology resulted in their quick development. Such methods include searching for common motifs within a family, or identifying conserved amino acid residues (e.g. Multiple Alignment – ClustalX-<http://www.clustal.org/>, [25]). An extraordinarily useful tool for the detection of distant homologues and rating proteins as belonging to the same family, is the PSI-BLAST algorithm and cascade PSI-BLAST [2,3].

TABLE 1. Commonly used databases for predicting gene function on the basis of homology.

DATABASE FUNCTION	DATABASE NAME AND WEBPAGE
Gene identification	EasyGene http://servers.binf.ku.dk/cgi-bin/easygene/search
	GenomeScan http://genes.mit.edu/genomescan/
Protein domains and motifs database	InterPro http://www.ebi.ac.uk/interpro/
	ProDom http://prodom.prabi.fr/prodom/current/html/home.php
Functional annotation	SWISS-PROT i TrEMBL http://expasy.org/sprot/

GENE INACTIVATION METHOD

The method of gene inactivation comprises of finding it and blocking transcription, which allows for a comparison of the obtained phenotype of the studied organism with the phenotype of the non-mutated organism. On this basis we can determine what changes have occurred in an organism and attribute them to the non-active gene. Currently, there are numerous methods of gene silencing used [43]. A basic principle of this technique is to generate and introduce a gene construction into an organism that will effectively block a specific gene. The effect

is the lack of synthesis of the protein encoded by the silenced gene, which often result in phenotypic differences that can lead to a conclusion concerning the function of a given gene.

One of the most common technique of gene inactivation is its discontinuation by means of an artificially introduced DNA fragment through insertional mutagenesis (knock-out) (Fig. 1), which is based on homological recombination. This technique is generally performed in one-cell organisms to avoid generating chimeras, whereby an organism is comprised of a mixture of mutated and non-mutated cells. Insertional mutagenesis is based on an insertion of the DNA fragment from a vector within the gene located on a chromosome. The chromosomal DNA obtained in this way contains the discontinued gene, which does not undergo expression and, in effect, there is no protein created. The disorders caused by the lack of protein show its function in the organism. Gene knock-out allows tracking phenotypical changes resulting from the exchange of sequence fragments between chromosomal DNA and the vector. The vector usually contains a gene discontinued by a selective marker allowing for an identification of recombinants and at the same time causing an interruption of the gene and making it inactive.

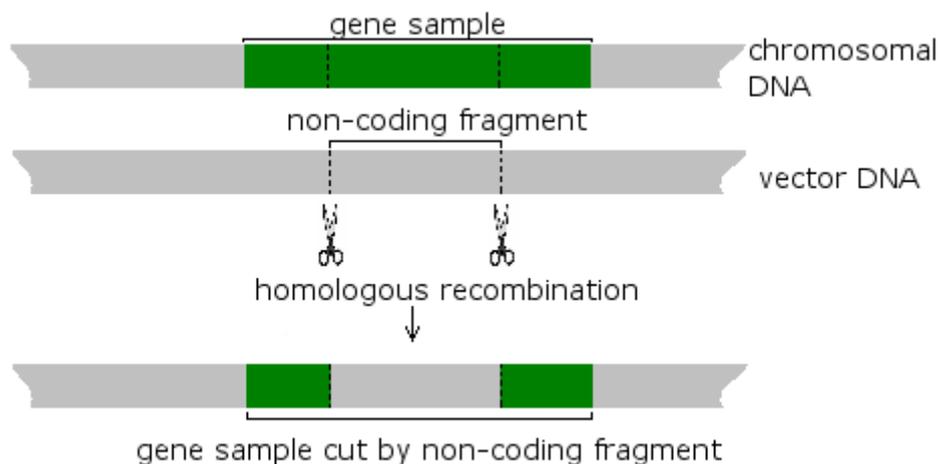


FIGURE 1. Gene inactivation by homologous recombination.

The differentiation between mutated and non-mutated cells is not always possible solely on the basis of phenotypical changes, which may be unequivocal. Thus, a so-called deletion cassette is often used to turn the studied gene off. The deletion cassette comprises of a selective marker, which is usually an antibiotic resistance gene, promoter sequences and two restriction enzymes sites (Fig. 2). Inserted into a cell in a vector, the deletion cassette is integrated into the host chromosomal DNA via homologous recombination. Using this technique, cells

containing the construct can be selected for due to the acquisition of antibiotic resistance (the most often used is kanamycin), as untransformed cells are non-resistant and accordingly die.

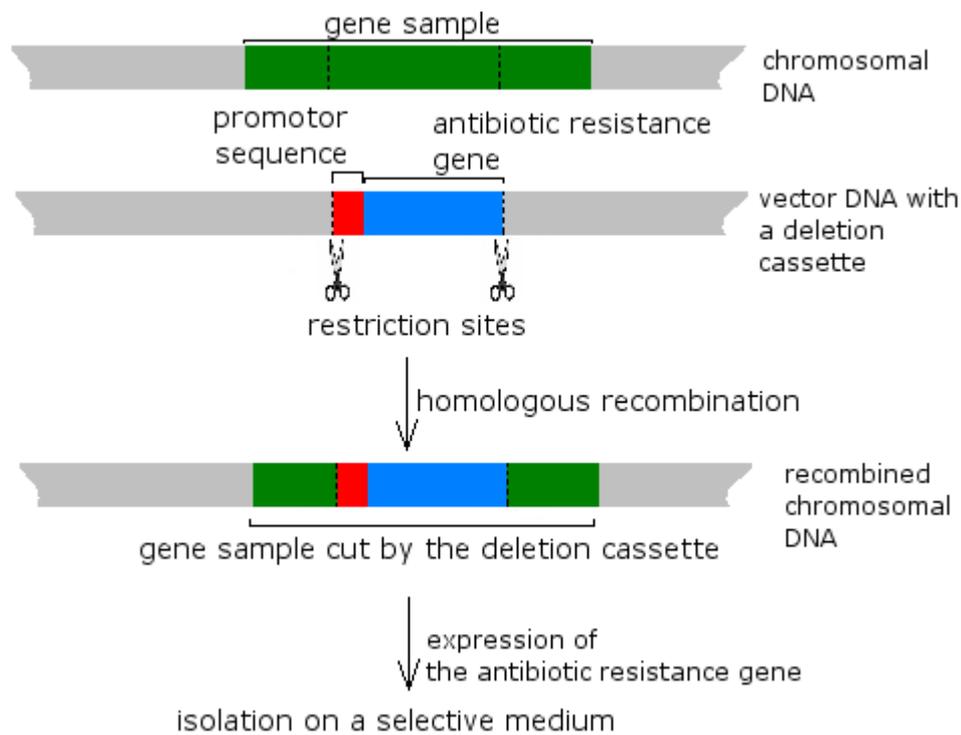


FIGURE 2. Gene inactivation by deletion cassette.

However, homologous recombination cannot be used in all cases. In a situation where the studied object is DNA of a multicellular organism, a single cell does not supply enough information about the influence of inactivation on the phenotype. The most commonly used multicellular model organism is the mouse, the genome of which is similar to humans. In order to avoid generating chimeras, embryonic stem cells are used. As these cells are totipotent, they can differentiate into any cell type and can effectively generate a whole organism. Unfortunately, as the inactivation of some genes is lethal, the progeny of healthy and mutated mice are used, in the hope that a phenotypic differences can be observed in the heterozygous state. To determine the function of genes that are embryo lethal, we use a system based on a tissue-specific Cre recombinase which recognises specific DNA fragments (loxP sequences). A mouse having the active gene guarded with loxP sequences is constructed. Then, a mouse with a modified gene is crossed with a transgenic mouse for Cre recombinase. The enzyme, being under the control of

the tissue-specific promoter undergoes transcription only in the specified tissue or organ and only in this place, the gene guarded with specific loxP sequences is 'knocked out'. Gene knock-out mice have been extensively used in medical studies to generate models of various diseases, such as cystic fibrosis, Duchenne muscular dystrophy or hemophilia A [20].

Conventional methods of gene silencing, including techniques based on homologous recombination are very time-consuming, meet numerous obstacles and require huge populations of transformants [11]. Expecting a 99% probability of finding one mutated gene (1 Kb in length) in *Arabidopsis thaliana* (genome size 120 Mbp) one would have to track 550 thousand independent transformants [23]. Currently, the most commonly used techniques are those based on posttranscriptional gene silencing, which comprise of a specific degradation of the produced RNA (RNA interference, RNAi) [30, 35, 41]. RNAi naturally occurs in both animals and plants and has been described by a joint notion of gene expression silencing with RNA. RNAi relies on the insertion of a double-stranded RNA molecule (dsRNA), which is cleaved into short interfering RNA (siRNA) by the RNaseIII enzyme Dicer. A siRNA duplex then becomes incorporated into a RISC complex which subsequently recognizes and degrades target complementary mRNA [6]. To achieve this purpose we also use the antisense sequential suppression (amiRNA – artificial microRNA), creating a complex with endogenous RNA and dsDNA resulting from this process is degraded. It was proved, however, that the insertion of sense sequence (co-suppression) is more effective and leads to the achievement of a higher level of silencing [43].

The gene inactivation with RNA has a huge potential as a tool in functional genomics. It allows studying the functions of genes in polyploidal organisms in which the use of such techniques as insertional mutagenesis is impossible [4].

THE METHOD OF INCREASED GENE EXPRESSION

The method of increased gene expression is to some extent a reverse of the gene inactivation method. Here we evaluate the phenotype caused by over-expression of a specific gene of interest and based on the comparison with the phenotype of the non-changed organism we can make a conclusion on the function of the studied gene.

The technique of increasing gene expression generally involves placing a multi-copy vector in the cells of the host (usually a model organism). The vector may have from 40 to 200 copies of the studied gene under the control of an active promoter, which contributes to the creation of a large amount of mRNA, and subsequently protein (Fig. 3).

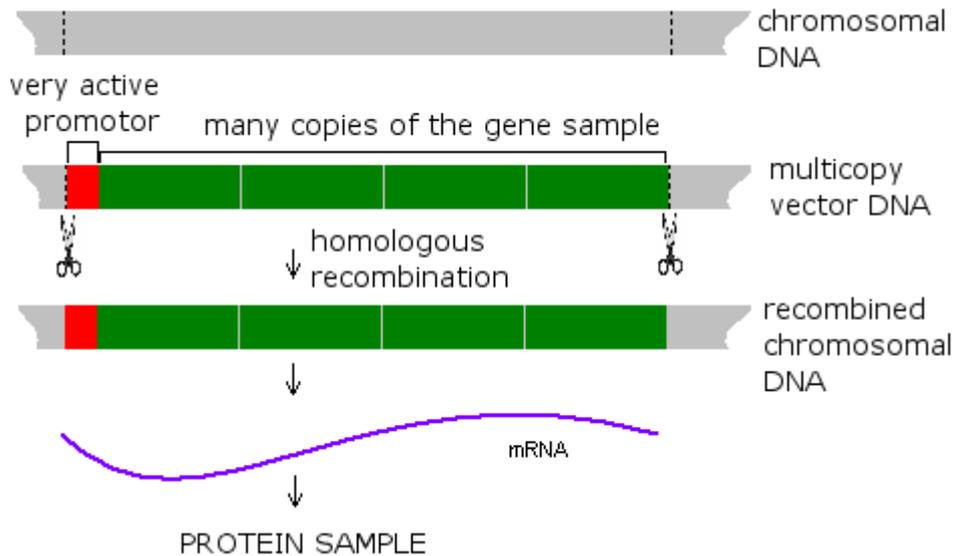


FIGURE 3. A schematic representation of gene overexpression.

The conclusions formulated on the basis overexpression studies performed with the use of this technique are sometimes inaccurate or unequivocal, as the reasons for phenotype changes in transgenic organisms are different [11]. The differences between changed and control organisms may result from either an increase or ectopic in gene expression. The problems that may be encountered while using this technique have persuaded the researchers to limit its use in favour of more modern and precise methods.

At the present time, overexpression studies are successfully used to produce recombinant proteins for biotechnology, pharmacy and medicine. The systems of recombinant proteins expression allow for a quick, inexpensive and safe creation of large amounts of the product. The most commonly used expression system is that of *Escherichia coli* due to the well-known genome and defined process of protein creation. Using this system, a series of proteins have been generated for use in medicine, such as the growth hormone as interferon, as well as most enzymes used in the laboratories of molecular biology [34, 40].

THE USE OF MICROARRAYS AND DNA CHIPS

Microarrays technology has been used to study gene expression for more than a dozen years. Undoubtedly, the advantage of a microarray is the possibility of

simultaneously studying many samples (even up to several thousand genes in one experiment) and comparing the profiles of their expression by means of clusterisation and an analysis of the main components (PCA). The advantage of PCA over clusterisation is visualisation in the form of 2D and 3D diagrams. It allows not only for an evaluation of the relation between the components, as it happens in dendrograms, but also for an evaluation of relations' degree between the elements in the space [32] (Fig. 4). The ongoing progress in the field of bioinformatics allows for a thorough statistical analysis of the obtained results. An analysis of microarray data uses tools and programs created in order to automatically analyse data on the basis of gene ontology, such as GOToolBox [29]. Gene ontology finds and describes the functions of genes, using the principles of hierarchisation and functional categorisation. It describes the characteristics of genes and their products in a given organism, cataloguing data into three areas of biological studies. These levels allow describing gene function, involvement in metabolic pathways and cellular localisation. The data processed in that way informs not only about the degree of gene expression, but also allows for a precise specification of the process time and place (tissue or organelle) that it takes place in. These are valuable pieces of information, since on the basis of their comparison with the genes already characterised by methods described earlier it can be concluded that gene transcripts of similar localisation and expression profile have a comparable function or take part in a commonly regulated process [27].

The principle of microarrays is very simple. On a glass or synthetic plate we put probes complementary with transcripts the expression of which we want to study. mRNA hybridizes specifically with the probe and the complex is made visible in a scanner measuring the level of fluorescence of earlier marked mRNA.

Commonly used are two types of microarray which differ in terms of origin and probe type. The probe may be a 0.5-2.0 Kb cDNA fragment (a DNA strand created as a result of reverse transcription on the RNA matrix) or a short oligonucleotide of 10-25 bp in length synthesised *in situ* on a plate. The microarrays with probes of the second type are called "DNA chips" [37].

Analysis of DNA microarrays and chip data is related to the method of predicting functions through homology as we can also use information collected in databases. Currently, there has been a dynamical growth of software packages available on the Internet, which combines the function of data analysis and storage in databases. The programs available on the market are created by commercial companies, university laboratories, as well as leading institutes in the field of genomics (Table 2). As a result of such variety we can properly choose software for the conducted studies [27].

Microarrays have been successfully used in functional genomics to study genes from fully and partially sequenced genomes [33]. Novel uses for microarrays are constantly being discovered and the high reliability of the microarray data has caused a rapid growth in many fields of scientific research.

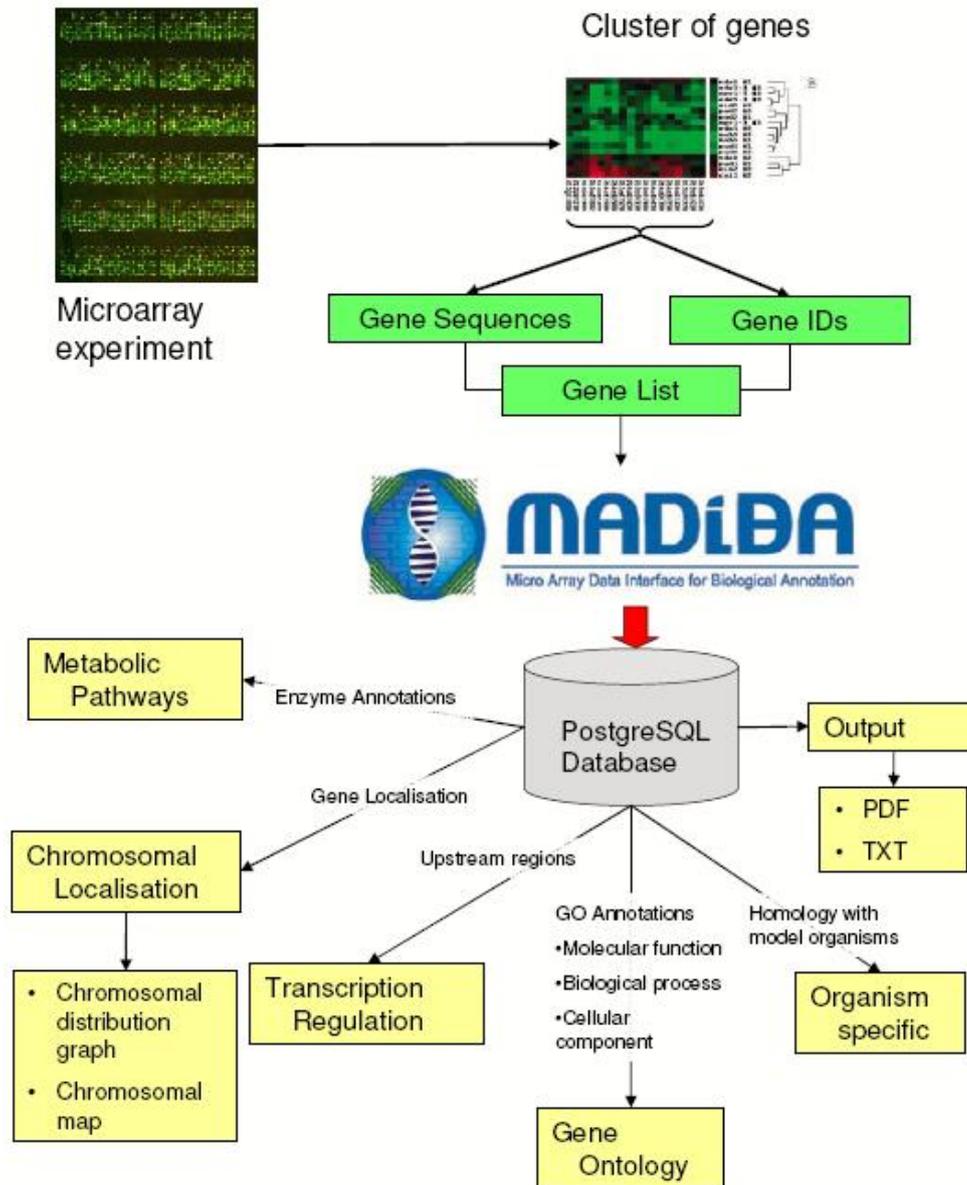


FIGURE 4. A schematic representation of the Madiba programme [according to 26].

The experimental systems, which use the interference of RNA and the principles of the microarrays method, were implemented into improving high-throughput techniques of functional analysis. The technique of transforming cells on plates (transfected-cell array- TCA) discovered recently seems to be a breakthrough in the large scale research on gene function. The technique combines the technology of microarrays with the methods of cellular biology, that is, transfection and cellular cultures. In the method of gene silencing with the use of TCA technique fragments of siRNA combined with transfecting agents are put on plates, while the whole thing is suspended in gelatine. Such matrix is covered with a plate with one layer of cells, which will undergo reverse transfection. Reverse transfection means introducing nucleic acid into a cell from the directly adjacent plate surface, which allows for a spatial separation of simultaneous transfection processes taking place on the plate. The effectiveness of transfection is evaluated on the basis of expression or silencing of a given gene, and the visualisation is possible through immunofluorescence. The large scale technique presented by Vanhecke and Janitz [42] allows for the analyses of more than 1,000 samples in a single experiment.

Medical experiments using microarray technology has enabled the development of diagnostic methods of detecting many diseases based on the recognised gene expression profiles in healthy and cancerous tissue [18]. In pharmaceutical studies, the technique is used to evaluate the safety and effectiveness of medicines prior to introduction into the market [9, 37]. It was also found useful in the industry for the production of plants with a higher expression of genes encoding phenotypical features beneficial from a customer's point of view [15].

THE USE OF TECHNIQUES ENRICHED WITH CHROMATIN IMMUNOPRECIPITATION (CHIP)

Genome-wide chromatin immunoprecipitation is a widely used technique enabling to track the interactions between proteins and DNA. These interactions fulfill an important role in the regulation of gene expression, maintaining the structure of chromatin or integrity of the genome. ChIP allows for the identification of transcription factor binding sites and locations of histones modification. The principle of this technique relies on treating the cells with a cross-linking agent (the most often used is formaldehyde), that generates covalent bonds between a specific protein and associated DNA. Cells are then lysed, genomic DNA is isolated and then sonicated in order to shear the chromatin into short fragments. During the next step, an antibody against the protein of interest is used to 'pull down' the protein of

TABLE 2. Commercially available programs used to analyse microarray data.

NAME	DESCRIPTION
Bioconductor	Software for statistical analysis of microarray data. Provides multiple tools for statistical and graphical interpretation of data [14].
MADIBA (MicroArray Data Interface for Biological Annotation)	MADIBA assigns biological meaning to clusters illustrating gene expression. Allows analysis of data drawn from clusters, visualization of metabolic pathways where the genes are implicated and defining common transcriptional regulatory factors of examined genes (Fig. 5) [26].
AMEN (Annotation, Mapping, Expression and Network)	AMEN enables detailed genomic research, analysing protein-protein interaction, chromosomal mapping, expression profiling and other proteomic data. Current version provides modules for uploading and pre-processing data from microarray expression profiling experiments, detecting groups of significantly co-expressed genes and searching database for gene function [8].
PGMapper	PGMapper is a software tool for automatically matching phenotype to genes from a defined genome region or a group of given genes by combining the mapping information from Ensembl database and gene function information from OMIM and PubMed database [44].
Mfuzz	Mfuzz is a software package used in the analysis of clustered data drawn from microarray experiments [24].
GEDI (Gene Expression Data Interpreter)	GEDI is a useful tool for large scale gene expression analysis in biomedical research [13].
GSE (Genomic Spatial Event)	Software for analysing gene expression data, genomic and functional annotations, results of algorithms. It is used to store and analyze all types of microarray data. Software has its own database [10].

interest and bound DNA. Cross-linking is then reversed, releasing the DNA from the protein interest [38].

Depending on further use, DNA fragments are usually amplified by PCR technique and then used as a template for either ChIP-chip or ChIP-seq.

The ChIP-chip technique, relies microarray technology to hybridise the obtained DNA fragments onto a microarray plate. A frequently underlined flaw of this technique is the accuracy of non-specific hybridization and background noise. Its also hard to create an appropriate wide-genome array. Most arrays have a limited number of spaces, thus they represent only a fraction of the whole genome. Another limitation of this technique may be the length of sequences located on the array, which reach several hundred base pairs. This makes identification of the binding site in the sequence difficult, especially for not well-recognised site of some transcript factors [21].

The ChIP-seq technique seems to be a more sensitive method yielding more reliable results than ChIP-chip [22]. Its development was made possible along with recent findings and developments in the high-throughput sequencing of DNA. The principle of this technique relies on the simultaneous sequencing of fragments obtained by ChIP, which are then mapped to a specific genome location. Commercial platforms and sequencing devices (e.g. SOLID System-AppliedBiosystems) make it possible to analyse 400 million short sequences on a single cycle [21]. The created genome maps with specific sites of binding transcription factors become a reliable basis for the identification of active, coding genome fragments. On the other hand, the knowledge concerning gene expression allows the scientists to manipulate transcription in a way to get to know the metabolic function of the product encoded by the analysed gene.

SUMMARY

Various methods, improved to minimise the risk of error, are used to determine gene functions. With time, the popularity of some techniques decreases in favour of others and completely new methods are developed, allowing for a more accurate study of uncharacterised genes.

The techniques using microarrays and chromatin immunoprecipitation are currently the most modern ones, and seem to be the most effective, mostly due to the relatively low costs (counting per number of simultaneously studied samples), a relatively high accuracy of results and short duration of the study. However, the databases used for comparing and concluding the functions of studied genes (also in case of the homology method) are largely based on the results obtained after the use of conventional methods, such as the inactivation method or gene

overexpression. Therefore, despite rapid technological developments, laboratories dealing with functional genomics have not completely abandoned these methods.

There is a high probability that thanks to this slowly progressing and often painstaking work there are going to be functions assigned to more and more uncharacterised genes which do not have homologues in databases. Many hypotheses have been suggested to explain the occurrence of ORFs, however, none of them is reliable enough to unambiguously explain the existence of this mysterious group of genes. It is suspected that the lack of sequence similarities to known genes results from the databases having too few homologues. It is also possible that ORFs have no homologues, as they may either have been lost by the ancestors in the process of evolution or are product of rapidly evolving genes, which makes it impossible to identify their initial sequences. The most intriguing hypothesis seems to be the one that ORFs may be non-coding RNA taking part in the regulation of expression or in the protection from parasites [28].

REFERENCES

- [1] AGI: THE ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000; **408**: 796-815
- [2] ALTSCHUL SF, GISH W, MILLER W, MYERS EW, LIPMAN DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403-410.
- [3] ALTSCHUL SF, MADDEN TL, SCHAFFER AA, ZHANG J, ZHANG Z, MILLER W, LIPMAN DJ. [Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.](#) *Nucleic Acids Res* 1997; **25**: 3389-3402.
- [4] AMBROS V, CHEN X. The regulation of genes and genomes by small RNAs. *Development* 2007; **134**: 1635-1641
- [5] BENSON DA, KARSCH-MIZRACHI I, LIPMAN DJ, OSTELL J, WHEELER LD. GenBank: update. *Nucleic Acids Res* 2005; **32**: 23-26
- [6] BERNSTEIN E, CAUDY AA, HAMMOND SM, HANNON GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 2001; **409**: 363-366.
- [7] BĄCZKOWSKI K, MACKIEWICZ K, KOWALCZUK M, BANASZAK J, CEBRAT S. Od sekwencji do funkcji – poszukiwanie genów i ich adnotacje. *Biotechnologia* 2005; **3**: 22-44
- [8] CHALMEL F, PRIMIG M. The Annotation, Mapping, Expression and Network (AMEN) suite of tools for molecular systems biology. *BMC Bioinformatics* 2008; **9**: 86.
- [9] CHAVAN P, JOSHI K, PATWARDHAN B. DNA Microarrays in Herbal Drug Research. *Evid Based Complement Alternat Med* 2006; **3**: 447-457.
- [10] DANFORD T, ROLFE A, GOFFORD D. GSE: A comprehensive database system for the representation, retrieval and analysis of microarray data. *Pacific Symposium on Biocomputing* 2008; **13**: 539-550.
- [11] FILIPECKI M, MALEPSZY S. Unintended consequences of plant transformation: a molecular insight. *J Appl Genet* 2006; **47**: 277-286
- [12] FUELLEN G. Homology and phylogeny and their automated inference. *Naturwissenschaften* 2008; **95**: 469-481
- [13] FUJITA A, SATO JR, FERREIRA CE, SOGAYAR MC. GEDI: a user-friendly toolbox for analysis of large-scale gene expression data. *BMC Bioinformatics* 2007; **8**: 457.
- [14] GENTLEMAN RC, CAREY VJ, BATES DM, BOLSTAD B, DETTLING M, DUDOIT S, ELLIS B, GAUTIER L, GE Y, GENTRY J, HORNIK K, HOTHORN T, HUBER W, IACUS S, IRIZARRY R, LEISCH F, LI CH, MAECHLER M, ROSSINI AJ, SAWITZKI G, SMITH C,

- SMYTH G, TIERNEY L, YANG JYH, ZHANG J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004; **5**: R80.
- [15] GIBSON G. Microarrays in ecology and evolution: a preview. *Molecular Ecology* 2002; **11**: 17-24
- [16] GINALSKI K, PAS J, WYRWICZ LS, GROTHUSS M, BUJNICKI JM, RYCHLEWSKI L. ORFeus: detection of distant homology using sequence profiles and predicted secondary structure. *Nucleic Acids Res* 2003; **31**: 3804-3807
- [17] GOGARTEN JP, OLENDZENSKI L. Orthologs, paralogs and genome comparisons. *Current Opinion in Genetics and Development* 1999; **9**: 630-636
- [18] GOLIAS CH, CHARALABOPOULOS A, STAGIKAS D, GIANNAKOPOULOS X, PESCHOS D, BATISTATOU A, SOFIKITIS N, CHARALABOPOULOS K. Molecular profiling and genomic microarrays in prostate cancer. *Exp Oncol* 2007; **29**: 82-84
- [19] GOWRI VS, SANDHYA S. Recent trends in remote homology detection: an Indian Medley. *Bioinformation* 2006; **3**: 94-96
- [20] GÓRSKA M, KOWALSKI ML. Knock-out genow- zastosowanie w badaniach medycznych. *Alergia, Astma, Immunologia*. 1997; **2**: 162-169
- [21] HOFFMAN BG, JONES SJM. Genome-wide identification of DNA-protein interactions using chromatin immunoprecipitation coupled with flow cell sequencing. *J Endocrinol* 2009; **201**: 1-13.
- [22] JOHNSON DS, MORTAZAVI A, MYERS RM, WOLD B. [Genome-wide mapping of in vivo protein-DNA interactions](#). *Science* 2007; **316**: 1497-1502.
- [23] KRYSAN PJ, YOUNG JC, SUSSMAN MR. T-DNA as an insertional mutagen in *Arabidopsis*. *Plant Cell* 1999; **11**: 2283-2290
- [24] KUMAR L, FUTSCHIK M. Mfuzz: A software package for soft clustering of microarray data. *Bioinformation* 2007; **2**: 5-7.
- [25] LARKIN MA, BLACKSHIELDS G, BROWN NP, CHENNA R, MCGETTIGAN PA, MCWILLIAM H, VALENTIN F, WALLACE IM, WILM A, LOPEZ R, THOMPSON JD, GIBSON TJ, HIGGINS DG. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; **23**: 2947-2948.
- [26] LAW JP, CLAUDEL-RENARD C, JOUBERT F, LOUW AI, BERGER DK. MADIBA: [A web server toolkit for biological interpretation of *Plasmodium* and plant gene clusters](#). *BMC Genomics* 2008; **9**: 105.
- [27] MACIEJEWSKI H, KONARSKI Ł, JASIŃSKA A, DRATH M. Analiza danych z mikromacierzy DNA- metody, narzędzia. *Journal Edited by Medical College- Jagiellonian University* 2005; **1**: 129-132
- [28] MACKIEWICZ P, BĄCZKOWSKI K, SOBCZYŃSKI M, CEBRAT S. Zagadka sierocych otwartych ramek odczytu (ORF-anów). *Biotechnologia* 2005; **70**: 45-56.
- [29] MARTIN D, BRUN Ch, REMY E, MOUREN P, THIEFFRY D, JACQ B. GOToolBox: functional analysis of gene datasets based on gene ontology. *Genome Biol* 2004; **5**: R101.
- [30] MARTINEAU HM, PYRAH IT. Review of the Application of RNA interference technology in the pharmaceutical industry. *Toxicol Pathol* 2007; **35**: 327-336
- [31] MERCHANT S, PROCHNIK S, VALLON O, HARRIS E, KARPOWICZ S, WITMAN G, TERRY A, SALAMOV A, FRITZ-LAYLIN LK, MARECHAL-DROUARD L, MARSHALL W, LIANG-HU Q, NELSON D, SANDERFOOT A, SPALDING M, KAPITONOV V, REN Q, FERIS, LINDQUIST E, SHAPIRO H, LUCAS S, GRIMWOOD J, SCHMUTZ J, CHLAMYDOMONAS ANNOTATION TEAM, JGI ANNOTATION TEAM, GRIGORIEV I, ROKHSAR D, GROSSMAN A. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 2007; **318**: 245-251.
- [32] NISHIMURA K, ABE K, ISHIKAWA S, TSUTSUMI S, HIROTA K, ABURATANI H, HIROSE M. A PCA Based Method of Gene Expression Visual Analysis. *Genome Informat* 2003; **14**: 346-347.
- [33] NOWAKOWSKA JA. Detekcja ekspresji genów drzew leśnych za pomocą mikromacierzy DNA. *Sylwan* 2006; **4**: 33-43
- [34] NUC P, NUC K. Produkcja rekombinowanych białek w *Escherichia coli*. *Post Bioch* 2006; **52**: 448-456
- [35] PAPROCKA M, WOŁOSZYŃSKA M. Potranskrypcyjne wyciszanie genów u roślin. *Kosmos* 2004; **53**: 193- 200

- [36] POPTSOVA MS, GOGARTEN JP. The power of phylogenetic approaches to detect horizontally transferred genes. *BMC Evolutionary Biology* 2007; **7**:45
- [37] ROMAN I. Mikromacierze DNA- perspektywy wykorzystania w badaniach skuteczności i bezpieczeństwa stosowania leków. *Post Bioch* 2008; **54**: 107-115
- [38] SCHMIDT D, WILSON MD, SPYROU CH, BROWN GD, HADFIELD J, ODOM DT. ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods* 2009; **48** :240-248.
- [39] SJÖLANDER K. Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics* 2004; **20**: 170-179
- [40] SORENSEN HP, MORTENSEN KK. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol* 2005; **115**: 113-128
- [41] SZWEYKOWSKA-KULIŃSKA Z, SZARZYŃSKA B. Nagroda Nobla 2006 za fundamentalne odkrycia w regulacji ekspresji genów u *Eucariontów*. *Post Biol Kom* 2007; **34**: 3-13
- [42] VANHECKE D, JANITZ M. Functional genomics using high-throughput RNA interference. *Drug Discov Today* 2005; **10**: 205-212.
- [43] WIŚNIEWSKA A, FILIPECKI M. Wyciszanie genów jako strategia badania ich funkcji w roślinach. *Post Biol Kom* 2003; **30**: 339-358
- [44] XIONG Q, QIU Y, GU W. PGMapper: a web-based tool linking phenotype to genes. *Bioinformatics Advance Access* 2008; **10**: 1323.

Corresponding author: Dorota Sitnicka

Department of Biochemistry, Faculty of Agriculture and Biology, Warsaw University of Life Sciences-SGGW

159 Nowoursynowska Str., Building 37, 02-776 Warsaw

phone: +4822 593-25-59; 507-148-235,

e-mail: dorota_sitnicka@sggw.pl