UPTEC X 02 025
JUN 2002

ISSN 1401-2138

DERYA KÜÇÜKGÖL

Identification of marker genes for preoperative diagnosis of thyroid tumours using microarrays

Master's degree project

Molecular Biotechnology Programme
Uppsala University School of Engineering

UPTEC X 02 025	**Date of issue 2002-06**

Author

Derya Küçükgöl

Title (English)

Identification of marker genes for preoperative diagnosis of thyroid tumours using microarrays

Title (Swedish)

Abstract

The advent of the large-scale method cDNA microarrays for global gene expression analysis, has enabled researchers to characterise tumour phenotypes in a much broader manner than was previously possible. The present study shows that the combination of comprehensive gene expression profiling and various mathematical algorithms reveals distinct gene expression patterns for different types of thyroid tumours. Gene expression profiling on thyroid cancer may in the future contribute to diagnostic information by providing a genetic "fingerprint" for each single tumour and patient. The aim is to obtain a better understanding for the underlying genetic differences in thyroid tumours and improved "patient-tailored" treatment.

Keywords

microarray, thyroid, tumours, gene expression, biomarkers

Supervisors

Anders Isaksson
Uppsala universitet

Examiner

Monica Nistér
Uppsala universitet

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **29**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

Identification of marker genes for preoperative diagnosis of thyroid tumours using microarrays

Derya Küçükgöl

Sammanfattning

Målet med detta examensarbete är att identifiera markörgener specifika för en typ av sköldkörtelcancer genom en storskalig analys av genetiska förändringar hos olika former av sköldkörteltumörer.

För detta syfte användes ”cDNA microarrayer” som möjliggör att man nu kan analysera uttrycket av tusentals gener samtidigt i enskilda tumörer. En cDNA microarray (kallas även ibland för "DNA-chip") består av ett objektglas på vilket tusentals korta sekvenser av DNA har placerats. Varje sekvens representerar en specifik gen, och metoden går ut på att mäta uttrycket hos ett stort antal gener i en och samma analys. Då man mäter uttrycket av tusentals gener i flera olika tumörer skapas stora mängder data. Denna information analyseras med avancerade matematiska metoder, och de olika mönster som framträder speglar genaktiviten hos de olika tumörerna. Man får således fram en typisk "genprofil" för varje tumör.

Genprofilstudier på sköldkörtelcancer kan i framtiden komma att bidra med diagnostisk information genom att man får fram ett genetiskt "fingeravtryck" för varje enskild tumör och patient. Förhoppningen är att man därefter skall kunna anpassa behandlingen med ledning av dessa genprofiler, så att varje cancerpatient får en skräddarsydd behandling som är så effektiv som möjligt.

Examensarbete 20 p i civilingenjörsprogrammet Molekylär bioteknik
Uppsala universitet juni 2002

Contents

1. INTRODUCTION

It has been noticed that cancer patients with seemingly the same disease respond differently to the same treatment with an unpredictable clinical course [1]. Some patients may have recurrence of the disease and develop distant metastases within a few years, while others remain disease free for a very long time, reflecting molecular differences. One of the ultimate goals in cancer diagnostics is to be able to provide each patient with a specific treatment, circumventing any unnecessary suffering for the patient not to mention the great cost for the health-care system. This type of “patient-tailored” therapy relies heavily on the existence of reliable techniques that can be used for classification and diagnosis of tumours.

Traditional tumour classification is based on the morphological appearance of the tumour cells, assessed under a microscope by a pathologist. However, this method has its limitations when distinguishing cancer subtypes with overlapping morphological features. Even pathologists with extensive experience may find it difficult to tell the benign hyperplastic adenomas from the well-differentiated malignant carcinomas. The need for a more sensitive and less subjective method is obvious. Gene expression profiling has been proven to be a powerful alternative for tumour classification. Analysing the levels of mRNA transcripts obtained from patient samples may provide accurate information on which cancer form the patient is suffering from.

In this study, we used cDNA microarrays to measure mRNA levels in primary thyroid tumours, consisting of benign adenomas and malignant carcinomas, to find significantly up- and down-regulated marker genes to be used for diagnosis. In this way, we hope to identify a gene expression signature that can be used for prediction of thyroid cancer outcome and to avoid unnecessary thyroid surgery.

2. BACKGROUND

2.1 Thyroid

2.1.1 The physiology and function of normal thyroid

The thyroid is a small butterfly-shaped gland situated in the neck, wrapped around the trachea just below the Adam's apple (Fig.1).

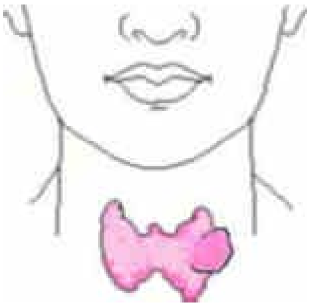


Figure 1. Thyroid with a nodule (Taken from www.endocrineweb.com/fna.html)

The thyroid is an endocrine organ and participates in a feedback mechanism. Upon stimulation from the hypothalamus through thyrotropin releasing hormone (TRH), the pituitary gland releases thyroid stimulating hormone (TSH). The thyroid reacts by releasing two hormones, triiodothyronine (T3) and thyroxine (T4), into the bloodstream. T3 and T4 control the metabolism and organ function of the body by means of converting oxygen and carbohydrates into energy. The thyroid gland traps the iodine supplied by food, iodized salt or other supplements to produce T3 and T4. The "3" and the "4" refer to the number of iodine molecules in each thyroid hormone molecule. Inadequate iodide intake as well as a variety of other factors, such as radiation treatment, can contribute to thyroid related diseases. Changes in hormone levels may cause serious disorders affecting among others heart rate, muscle strength, vision and mental state [2].

2.1.2 Thyroid Tumours

Epithelial tumours, including thyroid tumours, can roughly be divided into two groups; adenomas representing the benign tumours and carcinomas which are malignant tumours. Benign tumours are in general well-differentiated showing a similar morphology to the normal tissue while malignant tumours have deficient growth control. Other characteristics of benign tumours include cell uniformity, slow growth and few mitoses. Malignant tumours, on the other hand, show variation between the tumour cells, is locally invasive and have many mitoses. The ability to metastasize is what definitely distinguishes the malignant tumour from the benign.

However, the border between adenomas and carcinomas is not always obvious and can be rather diffuse, especially with the more traditional way of analysing tumours based on morphological tumour classification. Hence, diagnostic tools to analyse RNA expression levels and changes in the DNA can be used as a supplement to the morphological analysis.

2.1.3 Thyroid Cancer

Thyroid cancer is a fairly uncommon type of cancer and considered to be a "good" cancer in the sense that up to 95% of the affected patients survive the diseases. Thyroid cancer can appear as swelling or nodule in the thyroid gland. Other symptoms include hoarseness, difficulties in swallowing and breathing, neck pain and swollen lymph nodes. Worth noticing is that 95% of all thyroid nodules are benign, i.e. not cancerous [2, 3].

The causes of thyroid cancer is largely unknown but the fact that women seem to be far more susceptible to the disease, suggests hormonal status to be involved. It has also been observed that thyroid cancers occur more frequently in patients with a history of external neck irradiation during childhood for reasons such as enlarged tonsils, enlarged thymus glands and Hodgkin's disease. An increase in thyroid cancer frequency has also been noticed in connection with exposure to nuclear radiation [2].

2.1.4 Four types of thyroid cancer

Thyroid carcinoma is a malignant tumour growing as nodules in the thyroid gland. 80-90% of all thyroid carcinomas in humans are differentiated which means that they outwardly look like the normal tissue of the thyroid gland. Hence, the only way to confirm whether a nodule is malignant or not, is to take the biopsy of the specific nodule [4].

There are two main types of differentiated thyroid carcinoma; the *papillary* and *follicular thyroid carcinomas.* The papillary carcinoma can be identified by its nipple-like projections that can be observed when examined with a microscope. Papillary carcinomas usually grow slowly and metastasise locally to lymph nodes in the neck. Papillary tumours account for about 70% to 80% of all thyroid carcinomas, and can occur at any age, but the patients have good chances to recover if the disease is diagnosed and treated in time [2]. When the thyroid carcinoma is made up of small spherical structure called follicles, it is then called follicular. Follicular carcinomas account for 10-15% of the thyroid cancers and affects older patients more frequently. This form of cancer is considered to be more invasive than the papillary cancer and it can spread through the blood vessels to distant areas such as lungs and bones. However, due to their slow growth, the tumours can be detected and removed at an early stage and the prognosis is good [2].

Medullary thyroid carcinoma is the third most common type of thyroid cancer, and usually originates in the upper central lobe of the thyroid. It is more invasive compared to the papillary or follicular carcinomas and does not arise from thyroid hormone producing cells, but instead from C cells that make the hormone calcitonin. Medullary thyroid carcinoma can be hereditary but patients have a good prognosis [2].

Anaplastic thyroid carcinoma is the most malignant of all thyroid cancers. The tumour spreads primarily to the lungs, but also bone and brain metastases occur with high frequency. Due to its very aggressive nature, the patient often dies shortly after the diagnosis [2].

2.1.5 Treatment of the thyroid cancer patients

The most common way of eliminating the cancers that have not spread outside the thyroid gland is by surgically removing parts of the thyroid gland or the whole thyroid gland. The side effects of thyroidectomy include damage to parathyroid glands leading to temporary or permanent hypoparathyroidism that causes low levels of calcium in the bloodstream resulting for example in an increased excitability of nerves. Also temporary or permanent damage to the vocal cord may occur. [5, 6]

Following surgery the patients are offered regular follow-up examinations to detect recurrences and they need to administer thyroid hormone for the body to function normally [3]. Patients with papillary and follicular carcinomas go through radioiodine therapy (RAI), which involves intake of radioactive iodine. If there is any thyroid tumour tissue left, it will be killed by the absorbed radioactive iodine. RAI treatment can also be applied when the cancer has managed to spread to the surrounding tissues and to distant sites [3]. Medullary and anaplastic tumours do not accumulate the radioactive iodine and hence the patients are not subjected to RAI. Side effects of this therapy include nausea, inflammation of the salivary glands, vomiting, exhausting and bone marrow suppression [6]. External radiation is provided when total removal of the cancerous thyroid tissue is not possible [3].

Only 10% of the resected thyroid tumours are estimated to be carcinomas [7] and considering the unpleasant side effects for the patients, there is a strong interest in finding tools that can accurately discriminate between thyroid adenomas and carcinomas in an efficient way. Identifying biomarkers of malignant thyrocytes using microarrays seems to be a very attractive approach and is the main purpose of this project.

2.2 Microarray technology

The increasing amount of raw data produced by genome sequencing projects urges the development of more effective and faster genome-wide expression analysis, including differential display-PCR (DD-PCR), serial analysis of gene expression (SAGE), cDNA microarrays and oligonucleotide chips [8]. Compared to DD-PCR and SAGE, cDNA microarrays and oligonucleotide arrays have the capacity of revealing gene expression profiles of thousands of genes in a single experiment in a much faster and less expensive manner. Also, since the interaction between target and probe involves a huge amount of factors comprising the method for labeling target/probe, hybridization conditions, the gene sequence, etc., the array technology is better suited to measuring relative amounts of RNA from two samples versus absolute amounts [9].

The study of the thyroid cancers was conducted with cDNA microarrays, hence merely the cDNA microarray technology will be explained.

2.2.1 Microarray applications

In the medical field, microarrays can be used to diagnose the clinical outcome of various diseases including cancer by identifying marker genes with specificity for that particular disease. For this purpose it is not necessary to know the biological function of the genes since the aim is just to distinguish different classes and subclasses of the disease.

Functional genomics is another application area of microarray technology. Several genome-sequencing projects have now been completed, producing huge numbers of novel genes which we need assign function to. Today, this is mostly accomplished by searching for homologous proteins of known function. However, this method has its limitations, since genes with similar or even identical sequences may give rise to different proteins and, in reverse, proteins with the same function may derive from non-homologous sequences [10]. Gene annotation based on expression profiling builds on a darwinistic theory that genes are only expressed in certain cells under certain conditions when contributing to the overall fitness of the organism [9]. Hence, there may be a connection between the expression profile of a certain gene over a range of different conditions and its function.

Other applications of the microarray technology include identification of drug targets and tumour biology. The effect of drug metabolism or pathogen infections on the gene expression in cells can also be studied with microarrays [11].

2.2.2 cDNA microarrays

In cDNA microarrays, thousands of gene-specific DNA sequences derived from the 3'- or 5'-end of RNA transcripts (ESTs) are arrayed by a computer-controlled robot in microscopic amounts on solid substrates such as membranes or microscope slides coated with poly-lysine, amino silanes or amino-reactive silanes [12]. The DNA sequences are PCR-products, prepared in 96-well or 384-well plates, and derived from purchased cDNA libraries in which the cDNA fragments have been cloned into plasmid vectors. These vectors are transformed into *Eschericia coli* where it can be multiplied during the culturing process of the bacteria. The plasmids are then isolated and cDNAs are amplified with polymerase chain reaction (PCR) [13].

Total RNA or mRNA, isolated from test and reference cells, is reversely transcribed into cDNA and fluorescently labelled by the incorporation of Cy3 or Cy5-tagged nucleotides. These fluorescently labelled deoxyribonucleotides are spectrally distinct and can be incorporated into cDNA with high efficiency by the reverse transcriptase. After the labelling, the two cDNA samples are pooled together and applied on the array [13] (Fig. 2).

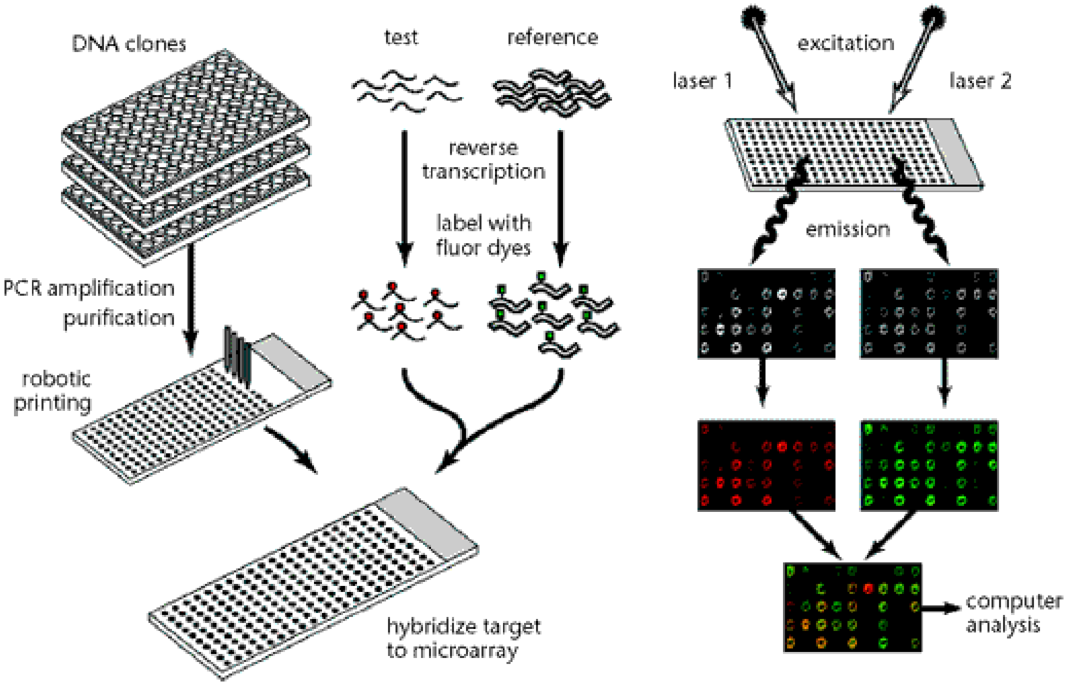


Figure 2. A schematic illustration of the cDNA procedure (Taken from Duggan et al, 1999)

2.2.3 Correlation between mRNA and protein abundance

For classification and diagnosis of diseases, there is no need to know whether the mRNA and protein levels correlate or not. However, other applications of microarray technology include predicting protein expression levels of a biological system from quantitative mRNA data. This assumes that there is a certain correlation between mRNA and protein levels, which is an issue of discussion. Several studies have been performed on this subject with various outcomes.

Gygi *et al.* [14] showed in one study that the correlation between yeast mRNA and protein levels is not good enough for prediction of protein levels from mRNA transcripts. In this study, the protein levels were estimated with two-dimensional gel electrophoresis (2DE) separation followed by scintillation counting of the selected proteins. The selected proteins were identified by mass spectrometry and protein database searching after tryptic digestion. The cognate mRNA transcript levels were obtained from SAGE frequency tables. For the 106 selected genes, a general trend of increased mRNA transcript levels resulting in increased protein levels could be noticed. Gygi *et al.* used the Pearson product-moment correlation coefficient (r_p) to measure the correlation between mRNA and protein levels. For the 106 genes, r_p was estimated to 0.94. However, this value is considered to be misleading since a few highly abundant proteins bias it. A new r_p for a subset of data, including genes with only 10 copies/cell and representing 69% (73 of 106) of data, was estimated to 0.356. Gygi *et al.* also observed that the mRNA transcript levels of proteins of similar abundance varied as much as 20-fold and inversely for proteins of similar quantity the mRNA levels differed by as much as 30-fold. Gygi et al point out that the bad correlation between mRNA and protein expression levels is an indication of posttranslational mechanisms controlling gene expression, hence mRNA expression profiles by themselves are not enough for quantitative analysis of biological systems.

In another similar study by Futcher *et al.* [15], they came to the opposite conclusion. Futcher *et al.* claimed that "considering the wide range of mRNA and protein abundance" there is a good correlation between mRNA and protein levels. Furthermore, they believe that Gygi *et al.* might have been mislead by using Pearson product-moment correlation coefficient, since it is valid only if both mRNA and protein levels were normally distributed. Instead they used Spearman rank correlation coefficient that can be applied on data, which is not normally distributed.

From the studies above, one can conclude that, although not strictly, there is a correlation between gene expression levels and protein expression levels. In the future it seem inevitable that there will be more use of proteome chips, to obtain a more complete picture of what is going on in the cell.

2.3 Data analysis

The way from the hybridised array to identifying differentially expressed genes is long and includes several steps including slide scanning and image analysis followed by data normalisation and analysis.

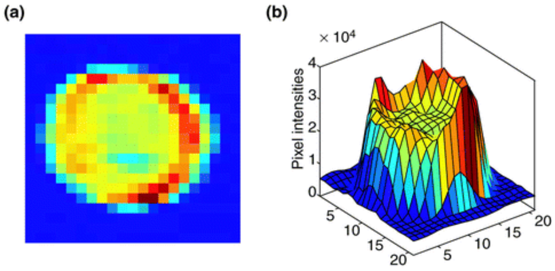
2.3.1 Slide scanning

The fluorescent image of the hybridised array is visualised using a confocal laser scanner with a photomultiplier tube (PMT) detector. For cDNA labelled with the Cy3-UTP and Cy5-UTP, a scanner with lasers generating light with a wavelength suitable for the excitation spectra of Cy3 (excitation wavelength of 532 nm) and Cy5 (excitation wavelength of 635nm) is used. The Cy3 and C5 channels are scanned separately and stored as TIFF files. These images can be merged and result in one single image with red and green spots representing over-expressed and under-expressed genes, assuming that the test sample is labelled with Cy5 and the reference sample is labelled with Cy3. Unchanged gene expression is visualised as yellow spots. These two images comprise the raw data from which relative expression levels of each gene can be calculated and differentially expressed genes can be identified.

2.3.2 Image Analysis

Next step in the microarray data analysis is the processing of the images obtained from the laser scanning.

The spots have to be defined and distinguished from the non-specific background signals that can arise due to various hybridisation artefacts or contaminants such as precipitated DNA fragments, fingerprints or dust on the surface of the slide. This is usually done with a grid that is adjusted to the array layout and the process can be both tedious and time consuming due to the irregular shape and size of the spots and uneven slide backgrounds. Both median and mean intensity of each spot, for red and green channels respectively, are available and can be used in the data analysis. Since the signal intensity is not uniformly distributed over the spot (Fig.3), choosing the median intensities seems to be a more robust method of capturing the amount of labelled cDNA on a certain spot [16].



TRENDS in Biotechnology

Figure 3. An illustration of the variation of pixel intensities over a hybridised spot (a) from the top and (b) from the side. The intensities range from low intensities (blue) to high intensities (red). (Taken from Hess *et al*, 2001)

Following identification of spots and its borders, the next step includes estimation of the background-subtracted fluorescence intensities of each spot. One approach is to estimate a global background based on the entire image. However this method does not take the uneven background, which arises during the hybridisation process, into account. An alternative approach, which was used in this project, is to determine the background of each spot locally, by measuring the mean or median intensities of a restricted area around the spot.

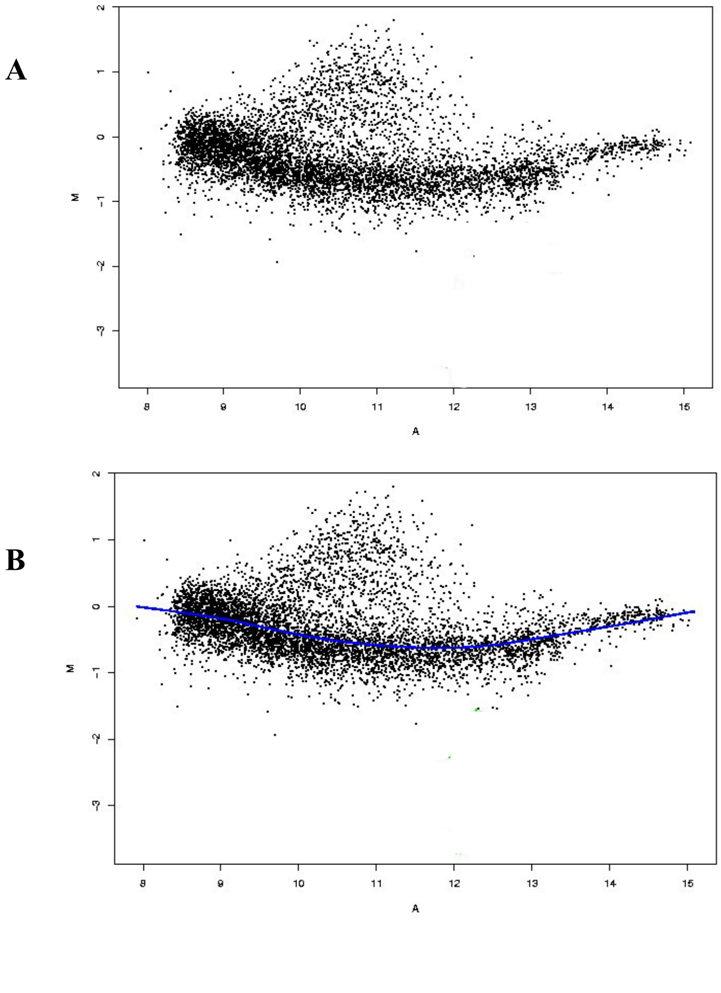
2.3.3 Normalisation

The raw data extracted during image processing includes a table with signal and background intensities, which must be further analysed in order to identify the differentially expressed genes. To remove systematic errors such as different incorporation of dyes or different amounts of RNA from the co-hybridised samples, and to be able to compare different slides with each other [17], the raw data must be normalised in terms of balancing red and green intensities. The data can be presented in a "MA" plot, where $M = \log_2 R/G$ is plotted versus $A = \log_2 \sqrt{(R*G)}$. R and G denote the intensities of the red and green channels. The reason to taking the logarithm of the ratio instead of the ratio itself is to achieve the same absolute value for a certain fold of up-

regulation or down-regulation. For example if a gene is up- or down regulated two fold, the ratios are 2 or 0.5 while the $\log_2$ values are +1 and –1, respectively.
Assuming that only a small fraction of the genes will be differentially expressed and the up- and down-regulation of the genes are symmetric and assuming that the amount of RNA hybridised on the array is the same for the co-hybridised samples, one can use the total intensities of the green and red channels and derive a normalisation factor with which one can bring the total intensities to same level.

Including so-called house keeping genes on the array can be another approach for normalisation. These genes are considered to have a constant expression in all conditions, hence should be present in equal amounts in the test and reference samples and yield similar red and green intensities. An alternative to housekeeping genes is to use control genes that are not related to the organism being studied. The DNA sequences of the control genes are spotted on the array and the corresponding mRNA is added at equal amounts to the test and reference sample, respectively. This should, just like in the case of housekeeping genes, result in spots with equal red and green intensities [18].

The global normalisation methods above are linear, meaning that a constant normalisation factor is applied for all genes to eliminate the systematic variations deriving from the differences in RNA amounts and dye intensities. Global normalisation methods are still extensively used, but “MA” plots have revealed an intensity dependent dye bias and hence non-linear algorithms such as the curve-smoother *lowess* is being preferred over global normalisation methods [18]. Spatial and print-tip effects on fluorescence intensities can also be corrected by normalisation (Fig. 4).



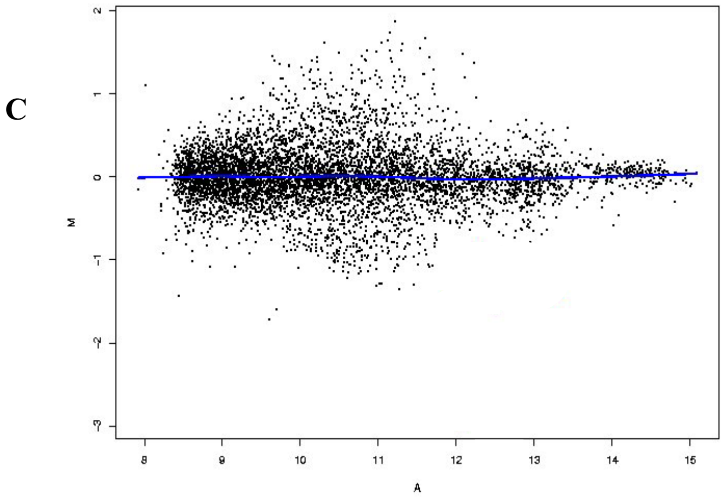


Figure 4. An example of an "MA" plot, before normalisation (A) and after lowess (B) and print-tip normalisation (C). (Taken from http://stat-www.berkeley.edu/users/terry/zarray/Talks/spie-norm/spie-norm.ppt, 2 Jun. 2002)

2.3.4 Methods for determining differentially expressed genes

There are a few methods of determining the differentially expressed genes, some of them less good than others. They are based on the idea that the more differentially expressed they are, the higher scores they receive.

A simple and crude way of ranking the differentially expressed genes in two cell samples is by measuring their mean M values. M_{ij} for gene *i* on array *j* is

$M_{ij} = \log_2 R_{ij}/G_{ij} = \log_2$ (expression level in test sample) $_{ij}$ / (expression level in reference sample) $_{ij.}$

R and G refer to the Cy5 and Cy3 fluorophor intensities that the samples are labelled with. High M values may result from differentially expressed genes, but it can also be derived from outliers [19] that have high M values due to for example hybridisation artefacts.

A more frequently used statistics to detect statistically significant expression ratios is the t-test, where for gene *i* the difference between group means is compared to the variability of the groups according to the formula

$t_j = (M_{Tj} – M_{Cj}) / \text{SQRT}[s^2_{Tj} / n_T + s^2_{Cj} / n_C].$

M_{Tj} and M_{Cj} denote the average expression level of gene *j* in the n_T test and n_C control microarray experiments, respectively. In the same way, s^2_{Tj} and s^2_{Cj} stand for variances of the gene *j* within the test and control groups. High absolute t_j value indicates that the gene *j* has statistically significant difference in its expression between the test and control hybridisations [17].

However, t-statistics has a tendency of generating false positive genes (i.e. genes that seem to be up-regulated when they are not), since small variances within the groups can lead to high t-values even though the means themselves are small. This error can be compensated for in a simple way by adding a constant to the standard deviation of each gene. For this purpose a statistical method called Significance Analysis of Microarrays (SAM) has been proposed by Tusher *et al*. [20].

2.3.5 Grouping

When grouping is unsupervised, it is called clustering. Clustering methods group genes based on the similarity without using any existing biological information. These methods suggests that if the function of one of the genes in a cluster is known, it can be hypothesised that the other genes are involved in the same biological process. Supervised grouping, also called classification, instead starts from a set of expression profiles for genes of known class label and learn to classify new profiles to one of the existing classes. Unlike the clusters produced by unsupervised learning, the quality of the predictions can be tested on another set of known genes, which were not used for learning [22].

3. MATERIALS AND METHODS

3.1 Preparation of labelled cDNA

3.1.1 Tissue samples

Tumour tissues were snap-frozen in liquid nitrogen and stored at –80ºC until use. The study was performed on two different classes of thyroid tumours with two patient samples from each. The benign class of tumours were constituted of the benign follicular adenoma, and the malignant class of follicular thyroid carcinomas. RNA derived from normal thyroid tissue was used as common reference.

3.1.2 RNA isolation and preparation

Life Technologies procedure using Trizol was followed to isolate total RNA from the tissues (for more details see *http://genomicscore.unc.edu/mRNA_Protocol_v2.html*). Frozen tumours were cut into smaller pieces of 50-100 mg each and homogenised in an appropriate volume of TRIzol Reagent (GibcoBRL Life Technologies, Maryland, USA). The TRIzol/tumour homogenate was centrifuged and any fat layer removed. 0.2 mls chloroform per ml of TRIzol Reagent was added to the TRIzol/tumour homogenate, which then was vortexed for 15 - 30 seconds and incubated at room temperature for 5 minutes. The sample was centrifuged at 12,000 g for 15 minutes at 4ºC and the resulting aqueous phase containing RNA was transferred to a fresh centrifuge tube without disturbing the interface. The RNA was precipitated by adding 0.5 mls isopropanol per 1 ml TRIzol Reagent into the sample and incubated for at least 10 minutes at room temperature followed by centrifugation at 12,000 g for 10 minutes at 4ºC. The supernatant was carefully removed and the white RNA pellet was washed once with 75% ethanol using 1ml 75% ethanol per 1ml TRIzol used. The tubes were centrifuged at 7500 g for 5 minutes at 4ºC and the ethanol was removed. After air drying the pellet for 10-20 minutes at room temperature the RNA was resuspended in nuclease free water.

The quality of the total RNA was assessed with spectrophotometry and electrophoresis where bright 28S and 18S band could be seen. Also Bioanalyzer (Agilent Technologies, California, USA) was used to check for the RNA quality. For isolation of mRNA, Qiagen's Oligotex kit (Qiagen, Valencia, California) was used (for protocol see Appendix 1). The mRNA was stored at -80 C until cDNA synthesis and labelling.

3.1.3 cDNA synthesis and labelling

Reverse transcriptase copies the mRNA into cDNA using both random and oligo dT primers as starting points for addition of nucleotides. The oligo dT primer pairs with the polyA on the 3' end of the mRNA while the random primer attaches randomly, hence yielding mRNAs of varying length. cDNA was prepared by direct labelling, which involves incorporation of dUTP-Cy3 or dUTP-Cy5 instead of dUTP during synthesis of the cDNA. Thus, the final product is fluorescent because the sequence contains U's that each carries a dye molecule. This method is robust, but requires a relatively large amount (25-100 µg) of total RNA to produce a strong hybridisation signal.

The direct labelling process was based on a modified version of the CyScribe First-Strand cDNA Labelling Kit from Amersham Biosciences (for details see Appendix 2). Originally, approximately 25 µg of total RNA was used for cDNA synthesis with only oligo dT primers, since random primers contribute to cDNA synthesis of rRNA and tRNA and thus giving rise to high background during the hybridisation. However, the outcome was consistent high background and weak signals. On the other hand, applying cDNA synthesis on isolated mRNA with both random and oligo dT primers, resulted in low background and high signals. Hence, approximately 50 µg of mRNA was used for each labelling reaction.

Two hybridisations were performed for each tumour using dye swapping, which means that the tumour sample is first labelled with Cy5-dUTP and then with Cy3-dUTP to assess for dye bias. In this case, dye bias means that incorporation of Cy5-dUTP and Cy3-dUTP depends on the sequence of the cDNA produced, hence resulting in somewhat different ratios for some genes when labelling is reversed. Genes showing big differences in their ratios in a dye swap experiment are in this way excluded from further analysis.

3.2 Hybridisation

The microarray technology is based on the intrinsic ability of single stranded DNA/RNA molecules to form duplexes with complementary sequences. This form of binding is both exact and reproducible. Hybridisation occurs simultaneously with both differentially labelled control and sample cDNA preparations and is dependent on the A-T and G-C pairing in the DNA duplex.

Hybridisation is performed either at 65ºC or 42ºC (if formamide is included). General and more specific blocking elements are included in order to decrease the unspecific signals. Cot-1 DNA (GibcoBRL Life Technologies, Maryland, USA) to block human repetitive DNA, yeast tRNA (GibcoBRL Life Technologies, Maryland, USA) to act as blocker of non-specific hybridisation and poly dA (Amersham Biosciences) to block oligo dT, were used in our hybridisations.

3.2.1 Pre-hybridisation

Pre-hybridisation was performed in order to decrease the unspecific background fluorescence according to protocol (see Appendix 2). Pre-hybridisation took place immediately prior to hybridisation.

3.2.2 Hybridisation

The hybridisation is carried out under a sealed cover slip in a humid chamber. For hybridisation procedure see protocol (Appendix 2). It is important that the probe is evenly distributed under the cover slip and that no air bubbles are introduced.

3.2.3 Washing slides

Hybridisation components such as SSC and SDS have an innate fluorescence and must be removed from the slide surface before scanning in order to avoid diffuse background fluorescence. All washes were performed at room temperature according to protocol (see Appendix 2).

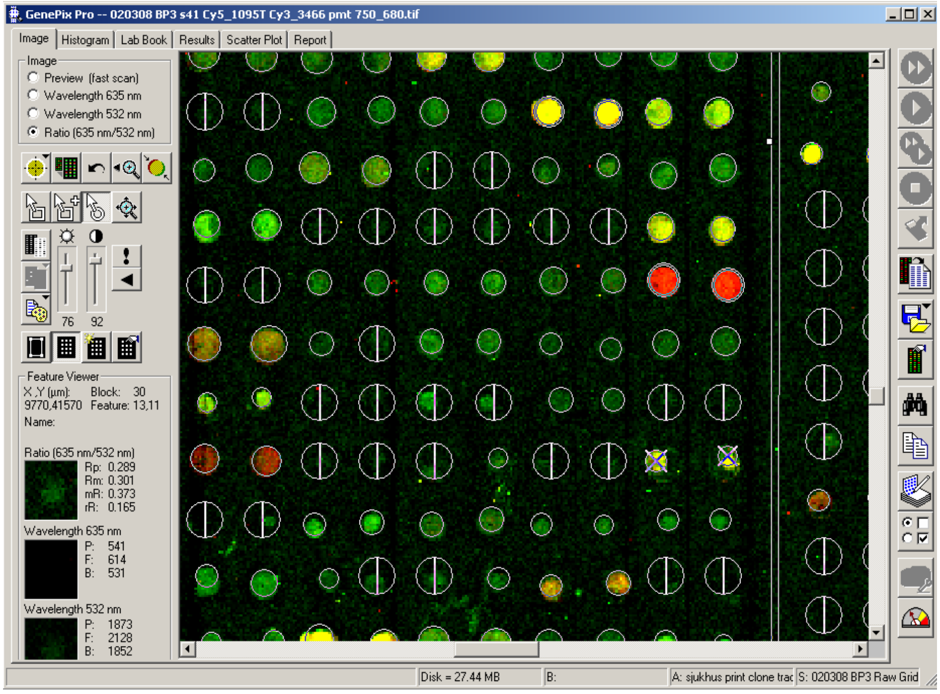
3.3 Data Analysis

3.3.1 Scanning

The fluorescent images of the hybridised microarrays were visualised by confocal laser scanning of the hybridised slide. GenePix 4000B (Axon Instruments, California, USA) was used. This instrument has two lasers for simultaneous signal detection in both the Cy3 and Cy5 channels and the obtained images are saved as greyscale TIFF files. The scanner was set to 10 micron resolution and the laser intensities were adjusted using PMT (Photo Multiplication Tube) to ensure that all levels of expression are detected and that the spot intensity is not saturated.

3.3.2 Image analysis

GenePix Pro 3.0 Microarray Analysis Software (Axon Instruments, California, USA) was used to extract data from the TIFF files generated from scanning the slide. This software possesses highly desired features such as auto-adjustment of spots within grids and alignment of non-overlaid images. A grid pattern is placed on the image to mark the location of each spot in the microarray. Good quality spots are easily defined by the auto-adjustment function while poor quality spots and artefacts are not as easy and are manually marked with the "flag" option and removed from further analysis. The ”flag” option enables us to define spots as good, bad, absent or not found, each of them associated with a value giving a hint about which spots should be discarded from further analysis (see Fig. 5).



GenePix Pro -- 020308 BP3 s41 Cy5_1095T Cy3_3466 pmt 750_680.tif

Image
○ Preview (fast scan)
○ Wavelength 635 nm
○ Wavelength 532 nm
● Ratio (635 nm/532 nm)

76 92

Feature Viewer
X,Y (µm): 9770,41570
Block: 30
Feature: 13,11
Name:

Ratio (635 nm/532 nm)
Rp: 0.289
Rm: 0.301
mR: 0.373
rR: 0.165

Wavelength 635 nm
P: 541
F: 614
B: 531

Wavelength 532 nm
P: 1873
F: 2128
B: 1852

Disk = 27.44 MB | B: | A: sjukhus print clone trac | S: 020308 BP3 Raw Grid

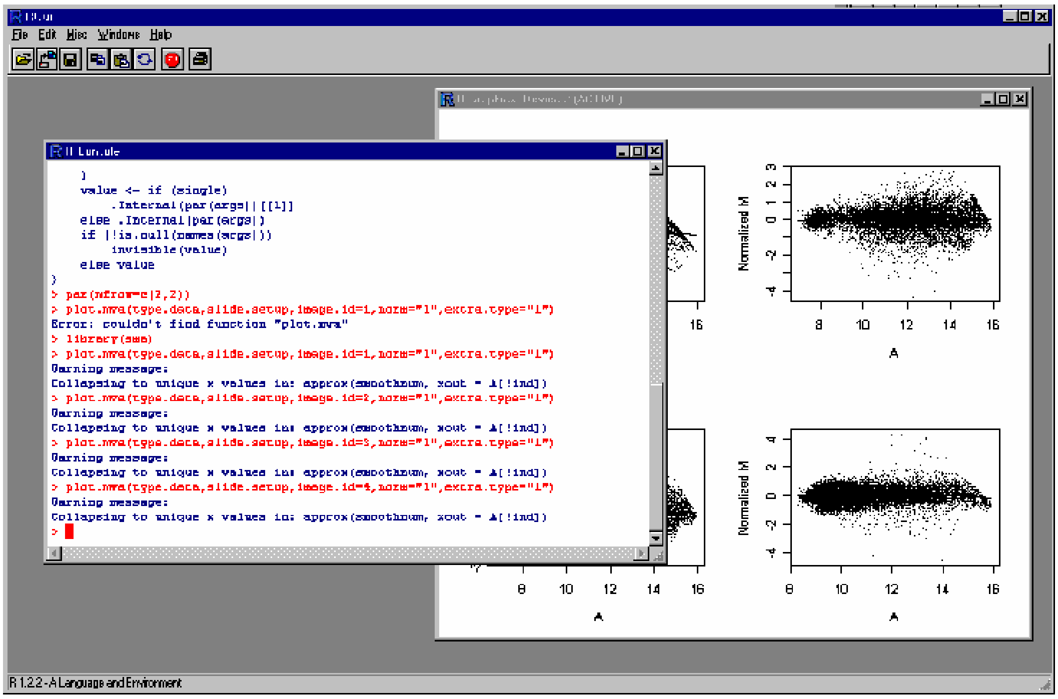
Figure 5. GenePix Pro 3.0 user interface.

GenePix Pro software measures simultaneously the Cy3 and Cy5 signal intensities while calculating a local background value for each spot. GenePix calculates the channel ratio for an individual spot from pixels with 5 different methods, i.e., ratio of medians, ratio of means, median of ratios, mean of ratios and regression ratios. These calculations convert the image file into a numerical data table of Cy3 signal, Cy3 background, Cy5 signal, and Cy5 background measurements.

In this study, the ratio is calculated from medians of each whole spot and median intensities were used since they are less subjected to extreme values [23]. The background-subtracted signal intensities were used for normalisation.

3.3.3 Normalisation

The raw data was normalised by using the robust scatter-plot smoother *lowess* function included in R, a freeware statistics language, which together with the SMA (Statistical Analysis of Microarrays-developed by Terry Speeds Microarray Data Analysis Group) can perform various microarray normalisations and data analysis. The normalised M and A values are then exported to an output file that can be opened in e.g. Excel.



R Gui
File Edit Misc Windows Help
R Graphics: Device 2 (ACTIVE)
R Console
Normalized M
A
8
10
12
14
16
-4
-2
0
2
4
1
3
R 1.2.2 - A Language and Environment

```
    }
    value <- if (single)
        .Internal(par(args))[[1]]
    else .Internal(par(args))
    if (!is.null(names(args)))
        invisible(value)
    else value
}
> par(mfrow=c(2,2))
> plot.mva(type.data,slide.setup,image.id=1,norm="l",extra.type="l")
Error: couldn't find function "plot.mva"
> library(sma)
> plot.mva(type.data,slide.setup,image.id=1,norm="l",extra.type="l")
Warning message:
Collapsing to unique x values in: approx(smoothdum, xout = A[!ind])
> plot.mva(type.data,slide.setup,image.id=2,norm="l",extra.type="l")
Warning message:
Collapsing to unique x values in: approx(smoothdum, xout = A[!ind])
> plot.mva(type.data,slide.setup,image.id=3,norm="l",extra.type="l")
Warning message:
Collapsing to unique x values in: approx(smoothdum, xout = A[!ind])
> plot.mva(type.data,slide.setup,image.id=4,norm="l",extra.type="l")
Warning message:
Collapsing to unique x values in: approx(smoothdum, xout = A[!ind])
> 
```

Figure 6. R user interface (version 1.3.1)

3.3.4 Data reduction

A filtering was performed to discard bad data and select for only “good” data from which the candidate genes were derived. Those genes, whose dye swap experiments differed from each other by more than 40% were excluded, even if it was only in one tumour sample. Only genes that managed to pass this criterion in all tumour samples was used to following steps of identifying candidate genes. This is a very stringent filtering, reducing the data dimensions greatly from 7700 genes to approximately 600 genes.

3.3.5 Candidate genes

For selection of a smaller subset of discriminative genes, two similar approaches were used, *SAM* and *BSS/WSS*. *SAM* (Significance Analysis of Microarrays) was used to identify the candidate genes. This is an Excel Add-In that correlates gene expression data to various clinical parameters such as treatment, diagnosis categories, survival time and time trends. This is accomplished by calculating a statistic $\boldsymbol{d_i}$ for each gene $\boldsymbol{i}$, indicative of the correlation strength. Repeated permutation of data is performed in order to correctly identify genes whose expression is significantly associated with the clinical parameter. A tuning parameter Δ **(**delta**)** is available to limit the false positive rate.

Another equivalent way of ranking the candidate genes was on the basis of the *BSS/WSS* ratio (between-class-variance/within-class-variance). However, like t-test this method runs the risk of generating false positive genes due to small variances within the classes leading to large *BSS/WSS* ratios. Hence, the subset of genes obtained from the *BSS/WSS* ranking was further evaluated by for each gene calculating the absolute difference between the mean values of M in adenomas and carcinomas, that is ABS (A*av*-C*av*) where A and C denote the average M values for adenoma and carcinoma tumours, respectively.

3.4 Microarrays

The microarrays were produced in house at WCN Expression Platform at Rudbeck laboratory, Uppsala University. The prepared probes were added to poly-L-lysine coated human cDNA glass arrays, containing 7700 different cDNAs (Invitrogen, Huntsville, USA) in duplicates. Microscope slides of size 75x25 mm were coated with Poly-L-lysine to give positively charged amine groups on the surface. The spotted DNA forms covalent bonds with the amine groups with the help of UV-crosslinking. The remaining free amine groups on the slides are blocked during the post-processing of the arrays with succinic anhydride, which is necessary in order to avoid unspecific background signals [24].

4. RESULTS

The aim of this project was expression profiling of thyroid tumours based on microarrays. The experimental design consisted of two adenomas and two carcinomas with their respective dye swap experiments. After processing and filtering of data, 15 most differentially expressed genes were extracted from BSS/WSS ratios and SAM. The expression levels of these 15 genes may in the future be used to discriminate the thyroid carcinomas from adenomas, when they have been confirmed in a larger clinical material.

4.1 Quality control

A quality control was conducted to assess if the quality of the slide was sufficient enough for further analysis and to see if any array is particularly bad and should fall out from the analysis. For each slide, the proportion of spots passing the criterion of signal-to-background ratio equal or greater than 2.5 was calculated. As it can be seen from Figure 7, the hybridised slides are of varying quality with percentage of spots managing this criterion, ranging from 12 to 34%. This means that out of 7700 genes, the expression values for between 900 and 2600 genes can be obtained. No experiment seems to be particularly bad, hence all the arrays were included in the further analysis. Also, it seems as none of the fluorophores show any significant differences in the incorporation efficiencies during the cDNA labelling.

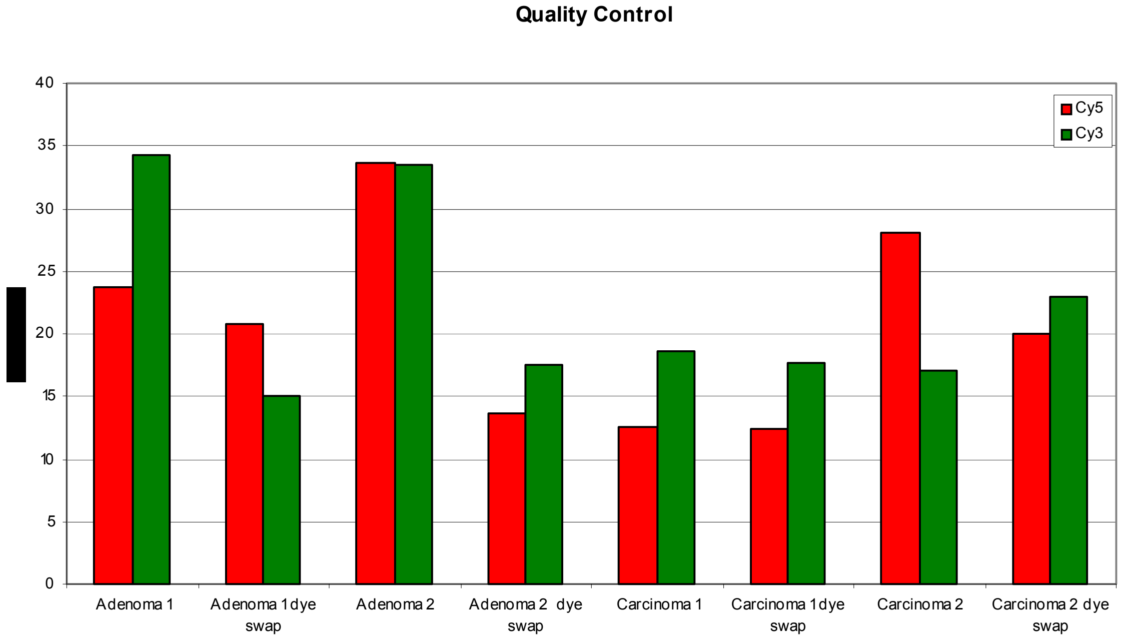


Figure 7. Quality control of the eight experiments. The columns represent the percentage of spots that have signal-to-background ratios of 2.5 or greater in the red and green channel, respectively.

4.2 Candidate genes

SAM and *BSS/WSS* ratios (see Materials and Methods, pages 19-20) were used to select for a smaller subset of candidate genes. From the SAM and *BSS/WSS* rankings, 15 most differentially expressed genes were selected. These 15 genes were ranked differently by these two methods but included in the top 30 genes with highest scores. The genes with their respective functions are presented in Table I.

Table I. Candidate genes identified by SAM and *BSS/WSS*

MSG1	Melanocyte-specific gene1(transcriptional activator in malignant melanoma)
CLDN1	Claudin 1 (tight junction protein)
SDR1	Short chain dehydrogen/reductase 1
IGFBP3	Insulin-like growth factor binding protein 3
IGFBP5	Insulin-like growth factor binding protein 5
ITGA3	Integrin alpha-3 subunit
DD96	Epithelial protein up-regulated in carcinoma, membrane associated protein 17
IGBP1	Immunoglobulin (CD79A) binding protein 1
CDC42	Cell division cycle 42 (GTP-binding protein, 25kD)
ARPC4	Actin related protein 2/3 complex, subunit 4 (20kD)
AKT2	V-akt murine thymoma viral oncogene homolog 2
BTAF1	BTAF1 RNA polymerase II, B-TFIID associated
RPS20	Ribosomal protein S20
SIAT9	Sialyltransferase 9 (CMP-NeuAc:lactosylceramide alpha-2,3-sialyltransferase; GM3 synthase)
CBS	Cystathionine-beta-synthase

To display how the candidate genes are expressed in the thyroid adenomas compared to the carcinomas, M ($\log_2$ (R/G)) values for each of the 15 candidate genes were plotted in Figure Y. The reddish staples present the M values in the carcinomas while the bluish staples displays the M values in the adenomas. In this diagram, one can distinguish which genes are up- or down-regulated compared to the reference sample consisting of normal thyroid tissue.

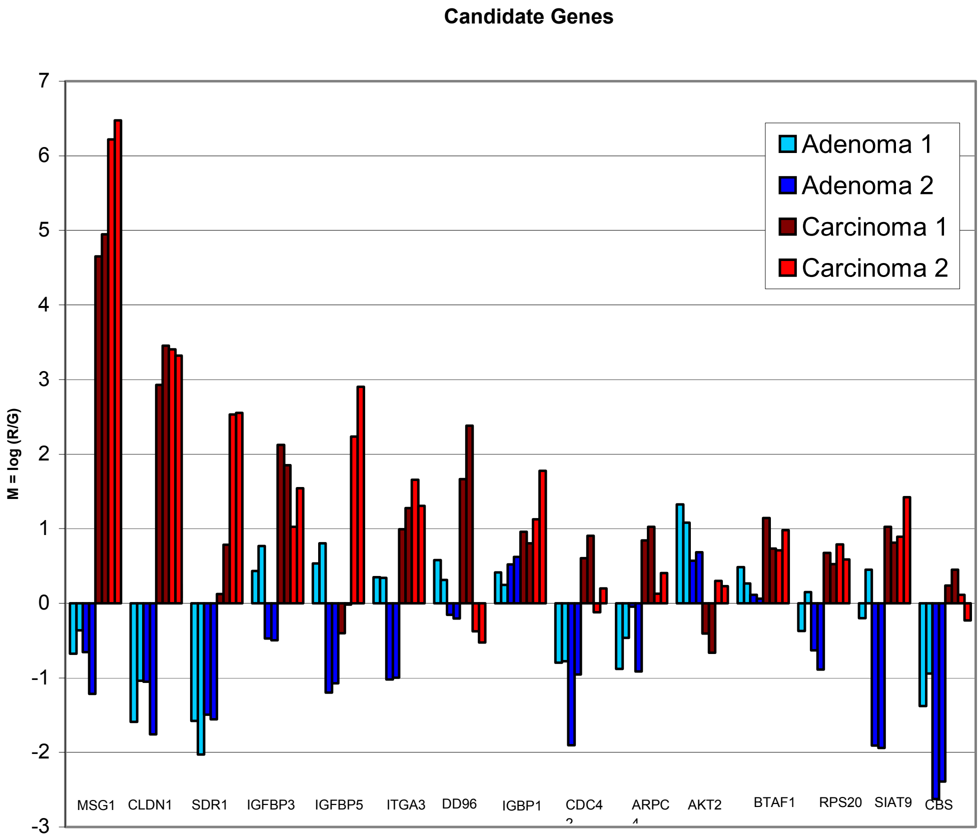


Figure 8. M(log$_2$ (R/G) values for the 15 candidate gene in adenomas (bluish staples) and in carcinomas (reddish staples), respectively.

MSG1 and CLDN1 are clearly the most differentially expressed genes between the adenomas and carcinomas, followed by SDR1, IGFBP3 and IGFBP5. IGFBP5 and DD96 were included as interesting genes since they are strongly up-regulated in one of the carcinomas and can be of importance for the classification.

5. DISCUSSION

5.1 All the experiments passed the quality control

The hybridisations were of varying quality according to our quality criterion that signal-to-background ratio should be 2.5 or greater. These variations are most likely due to differences in the hybridisation and cDNA labelling reactions. Also worth noting, is the fact that there is no consistent pattern of one fluorophore during the cDNA labelling. In general, the percentage of spots passing this quality criterion needs to be increased to improve our chances of finding good candidate genes. One way to do is to choose a slide with minimal autofluorescense contributing to the non-specific background signals. Applying a larger amount of labelled cDNA to microarrays may amplify the signals intensities.

5.2 Ranking differentially expressed genes with BSS/WSS and SAM

The filtering of data yielded a subset of about 600 genes whose dye swap experiments at most deviated by 40%. This may seem as a very crude data reduction but with so few tumour samples in each class. An alternative filtering approach could be tried to see if the same candidate genes can be obtained. The ranking lists from *BSS/WSS* ratios and SAM were used for selection of discriminative genes. These methods are based on similar algorithms and hence, not surprisingly, identify same candidate genes. SAM seems to be a better approach due to its in-built restriction for false positives. Worth noting is that none of these methods reveal the interactions between the genes. Other approaches based on a different discriminative algorithm should be used to verify the candidate genes.

5.3 Candidate tumour markers of thyroid carcinoma

The most discriminative genes seem to be involved in tumourigenesis of various cancer forms. MSG1 (melanocyte-specific gene1) is a transcription activator found to be over-expressed in malignant melanoma [25]. In a study by Ahmed *et al.* [25], the protein expression levels of MSG1 in human tumour samples from nevus and malignant melanoma were investigated using immunochemistry. Relatively high levels of MSG1 proteins could be detected in the melanoma while nevi samples show less extensive MSG1 protein expression. Also, in nevi the MSG1 proteins were restricted to pigmented regions while in melanomas MSG1 proteins could be seen in pigmented as well as non-pigmented areas. The results suggest that the MSG1 gene may be involved in pigmentation and malignant transformation of the pigment cells. In another study by Shioda *et al.* [26], it was found that MSG1 activates the transcription of Smad4 that has a key role in transforming growth factor-beta signalling.

In a recent study by Huang *et al.* [27], MSG1 was identified as a specific biomarker of papillary thyroid carcinoma (PTC). The protein product of MSG1 was studied by immunohistochemical staining to see if it correlated with the mRNA levels obtained from expression array and RT-PCR (real time-polymerase chain reaction) analysis. MSG1 proteins were found in 39/42 PTC, 0/6 follicular thyroid carcinoma, 0/1 anaplastic thyroid carcinoma, and 0/9 normal thyroid tissues. The result can be a bit puzzling and contradictory to our results derived from the follicular thyroid carcinomas. However, it is important to bear in mind, that the results obtained by Huang *et al* account only for the protein levels which does not need to be correlated with the mRNA levels as previously discussed. MSG1 seems to be an interesting gene and its expression in PTC should be studied along with the follicular thyroid carcinomas to see if we can come to the same conclusion as Huang *et al*.

The other candidate gene, CLDN1, is a tight junction protein, controlling cell-to-cell adhesion. It is also found to be involved in beta-catenin-Tcf/LEF signalling and frequently up-regulated in colorectal tumourigenesis [28]. Tight junction proteins are specific for epithelial cells, where it seals neighbouring cells together in an epithelial sheet to prevent leakage or molecules between them. Both thyroid and colorectal cancers are epithelial cancers, hence it makes sense that CLDN1 is strongly up-regulated in thyroid cancer.

Elevated levels of IGFBP3 have been showen to increase the risk for colon cancer [29]. However, it has also been shown that IGFBP3 has a tumour suppressive effect in prostate cancer and that IGFBP3 protein levels are strongly reduced in malignantant prostate cells [30] making it difficult to explain why IGFBP3 protein levels are significantly up-regulated in thyroid carcinomas. Different expression behaviour in different tumour types may explain its inconsistency.

5.5 Future objectives

In the future, more tumour samples should be included in the experimental setup to minimize the random variations in the selection of candidate genes and to be able to make classification of adenomas and carcinomas. It is also necessary to extract candidate genes from an independent test set of thyroid tumour samples to verify the differentially expressed genes obtained from the previous tumour sample set. Finally, RT-PCR analysis should be performed on these candidate genes to confirm that they are in fact differentially expressed. If RT-PCR confirms the microarray results, then two or three of the most differentially expressed genes could be used to discriminate the follicular thyroid carcinoma biopsies from the adenoma ones and right treatment could be given to the patient.

Microarrays will most likely not make its way to the clinics, not in the nearest future anyway. The technology is far too complicated and expensive. However, it can be used to quickly screen for e.g. the two most discriminative genes between two classes of disease cases. If we take follicular thyroid carcinomas to illustrate an example, it would be enough for the doctor to just take a biopsy of the thyroid tissue and run a RT-PCR on the tissue. If the MSG1 and CLN1 are up-regulated then it is likely that the patient has a malignant tumour and should go through a surgery. This would be a relatively quick and cheap screening method, providing a far less subjective classification then the one based on morphology.

Acknowledgements

First of all, I would like to thank everybody at the WCN Microarray Platform for giving me the opportunity to work with microarrays and making my time there so pleasant. I would like to address special thanks to my supervisor Dr. Anders Isaksson for taking the time to answer my questions and helping me to solve any encountered problems. Also, thanks to Mårten Fryknäs, for teaching me everything I know about microarrays plus some nice lab tricks. Finally, I would like to thank my examiner Prof. Monica Nistér.

List of Abbreviations

cDNA	Complementary deoxyribonucleic acid
DNA	Deoxyribonucleic acid
EDTA	Ethylenedinitrilo tetraacetic acid
HEPES	2-(4-(2-Hydroxyethyl)-1-piperazinyl)-ethanesulphonic acid
MRNA	messenger ribonucleic acid
PCR	Polymerease Chain Reaction
RNA	Ribonucleic acid
RT-PCR	Real Time- Polymerease Chain Reaction
SDS	Sodium dodecyl sulphate
SSC	Saline sodium citrate
TE	Tris-EDTA

REFERENCES

1. van't Veer LJ *et al*. (2002): Gene expression profiling predicts clinical outcome of breast cancer, *Nature* 415: 530-5
2. Thyroid Disease: http://thyroid.about.com (2 Jun. 2002)
3. The American Thyroid Association: http://www.thyroid.org (2 Jun. 2002)
4. EndocrineWeb.com: http://www.endocrineweb.com/fna.html (2 Jun. 2002)
5. Methodist Health Care System: http://www.methodisthealth.com/endocrin/hypopar.htm (2 Jun. 2002)
6. University of Pennsylvania Cancer Center: http://oncolink.upenn.edu (2 Jun. 2002)
7. Bartolazzi A *et al*. (2001): Application of an immunodiagnostic method for improving preoperative diagnosis of nodular thyroid lesions, *The Lancet* 357: 1644-9
8. Khan J *et al*. (1999): Expression profiling in cancer using cDNA microarrays, *Electrophoresis* 20: 223-9
9. Duggan DJ et al. (1999): Expression profiling using cDNA microarrays, Nature genetics supplement 21 (1 Suppl): 10-14
10. Noordewier MO and Warren PV (2001): Gene expression microarrays and the integration of biological knowledge, *Trends in Biotechnology* 19(10): 412-5
11. Debouck C and Goodfellow PN (1999): DNA microarrays in drug discovery and development, *Nature Genetics* 21(1 Suppl): 48-50
12. Brown PO and Botstein D (1999): Exploring the new world of the genome with DNA microarrays, *Nature Genetics* 21(1 Suppl): 33-7
13. Eisen MB and Brown PO (1999): DNA arrays for analysis of gene expression, *Methods Enzymol*. 303: 179-205
14. Gygi SP *et al*. (1999): Correlation between Protein and mRNA abundance in Yeast, *Molecular and Cellular Biology* 19(3): 1720-30
15. Futcher B *et al*. (1999): A sampling of the yeast proteome, *Molecular and Cellular Biology* 19(11): 7357-68
16. Hess KR *et al*. (2001): Microarrays: handling the deluge of data and extracting reliable information, *Trends in Biotechnology* 19(11): 463-8
17. Dudoit S *et al*, (2000): Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments, Technical report #578
18. Yang YH *et al*, (2002): Normalization for cDNA Microarray Data, SPIE BiOS, San Jose
19. Lonnstedt I and TP Speed, (2001): Replicated Microarray Data, Statistical Sinica, Accepted
20. Tusher *et al* (2001): Significance analysis of microarrays applied to the ionizing radiation response, *Proc Natl Acad Sci U S A* 98(9): 5116-21
21. Raychaudhuri S *et al* (2001): Basing microarray analysis: grouping and feature reduction, *Trends in Biotechnology* 19(5): 189-93
22. Hvidsten TR *et al* (2001): Predicting gene function from gene expressions and ontologies, *Pac Symp Biocomput.*: 299-310
23. GenePix user manual: http://www.axon.com/gpix_update/GenePix_Pro_3.0_User_Guide_Rev_B.pdf (2 Jun. 2002)
24. *Bio*Robotics Inc.: http://www.biorobotics.com/Users/downloads/BioNoteMG002_V1.0.pdf (2. Jun. 2002)
25. Ahmed NU *et al*. (2001): Aberrant expression of MSG1 transcriptional activator in human malignant melanoma in vivo, Pigment Cell Res. 14(2): 140-3
26. Shioda T *et al.* (1998): Transcriptional activating activity of Smad4: roles of SMAD hetero-oligomerization and enhancement by an associating transactivator, *Proc Natl Acad Sci U S A* 95(17): 9785-90
27. Huang Y *et al*. (2001): Gene expression in papillary thyroid carcinoma reveals highly consistent profiles, *Proc Natl Acad Sci USA* 98(26): 15044-9
28. Miwa N *et al* (2001): Involvement of claudin-1 in the beta-catenin/Tcf signaling pathway and its frequent upregulation in human colorectal cancers, *Oncol Res*. 12(11-12): 469-76
29. Palmqvist R et al (2002): Plasma insulin-like growth factor 1, insulin-like growth factor binding protein 3, and risk of colorectal cancer: a prospective study in northern Sweden, *Gut*. 50(5): 642-646
30. Dewi GR *et al* (2002): Insulin-like growth factor binding protein-3 induces early apoptosis in malignant prostate cancer cells and inhibits tumor formation in vivo, *Prostate* 51(2): 141-152

APPENDIX 1

Message RNA isolation with QIAGEN Oligotex kit Mini (Qiagen cat. no. 70022)

A. Reagents:

For mRNA isolation

- Binding Buffer **OBB**
  20 mM Tris-Cl pH7.5, 1 M NaCl, 2 mM EDTA, 0.2% SDS
- **Oligotex Suspension**
  10% (w/v) in 10 mM Tris-Cl pH7.5, 500 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.1% NaN_3
- Wash Buffer **OW2**
  10 mM Tris-Cl pH7.5, 150 mM NaCl, 1 mM EDTA
- Elution Buffer **OEB**
  5 mM Tris-Cl pH 7.5
- Small spin columns
  RNase-free spin columns

B. Procedure:

1. Dilute upto 250 µg total RNA in 250 µl ddH_2O.
2. Add 250 µl OBB (pre-heated at 70°C).
3. Add 12 µl Oligotex (pre-heated at 37°C).
4. Mix by pipetting up and down.
5. Incubate at 70°C for 3 minutes.
6. Incubate at room temperature for 10 minutes.
7. Spin at 14,000 x g for 2 minutes.
8. Remove supernatant and save for second extraction (step 20).
9. Resuspend pellet in 400 µl OW2 by pipetting up and down.
10. Transfer to spin column in clean 1.5 ml tube.
11. Spin at 14,000 x g for 1 minute. Discard flow-through.
12. Transfer spin column to clean 1.5ml tube.
13. Add 400 µl OW2.
14. Spin at 14,000 x g for 1 minute. Discard flow-through.
15. Transfer to spin column in clean 1.5ml tube.
16. Add 50 µl OEB (pre-heated at 70°C).
17. Resuspend Oligotex by pipetting up and down.
18. Spin 14,000 x g for 1 minute.
19. Repeat step 16 to18
20. Save eluted mRNA on ice.
21. Resuspend 7 µl Oligotex with supernatant from step 7 and transfer to 1.5ml tube. Repeat extraction from step 4. Combine eluted mRNA.

APPENDIX 2

Protocols for cDNA synthesis and hybridisation

Probe preparation by direct Cy-dNTP incorporation

A. Reagents:

mRNA
Oligo dT primer*
0.1M DTT*
CyScript buffer*
dUTP nucleotide mix*
Cy3-dUTP (Amersham Biosciences #53022)
Cy5-dUTP (Amersham Biosciences #55022)
Cy-script reverse transcriptase
2.5 M NaOH
2 M Hepes free acid
0.5M EDTA
TE
ddH$_2$O
Microcon YM-30 (Amicon #42410)

* in CyScribe First-Strand cDNA Labelling Kit (Amersham Biosciences # RPN6200)

B. Procedure:

1. Start with 50 µg total RNA resuspended in a total volume of 9 µl.
2. Add 2 µl of anchored oligo dT primer.
3. Heat to 70°C for 5 min in a water bath.
4. Let the mix cool to room temperature for 10 min.
5. Add the following to the tube:
   4 µl CyScript buffer
   2 µl 0.1M DTT
   1 µl nucleotide mix
   1 µl Cy-UTP (Cy-3dUTP or Cy-5dUTP)
   1 µl Cy-script reverse transcriptase
6. Mix and spin down in an eppendorf centrifuge at maximum speed.
7. Incubate at 42°C for 1.5 hours.
8. Stop reaction with 5 µl of 0.5 M EDTA.
9. Add 2 µl of 2.5 M NaOH.
10. Incubate at 37°C for 15 min.
11. Add 10 µl of Hepes free acid.
12. Mix well.
13. Mix the two differentially labeled RNA samples.
14. Load the mix on a Microcon YM-30 column.
15. Add 400 µl TE.

16. Centrifuge for 9 min at 14,000 x g.
17. Discard the flow-through.
18. Add 500 µl ddH_2O.
19. Centrifuge for 9 min at 14,000 x g.
20. Discard the flow-through.
21. Add 500 µl ddH_2O.
22. Centrifuge at 14,000 x g until 5-10 µl remains above the membrane.
23. Move the Microcon YM-30 column to a new tube.
24. Collect labeled cDNA by turning the Microcon YM-30 column upside down and spin for 1 min at14,000 x g.

Microarray hybridisation

A. Reagents:

Labeled cDNA
20XSSC
2.5% SDS
Denhardts solution
Yeast tRNA (20mg/ml) (Gibco/BRL # 15279-011)
Cot1 DNA (10mg/ml) (Gibco/BRL # 15401-029)
A_{80} oligo 0.85µg/µl
Washing solutions: 1. (1XSSC, 0.2% SDS), 2. (0.4XSSC), 3. (0.2XSSC)

B. Pre-hybridisation:

1. Mix: 25 µl 20XSSC
10 µl 50XDenhardt's solution
1 µl Yeast tRN
5 µl 10% SDS
50 µl 100% Formamide
9 µl H_2O
2. Add 20 µl of 3XSSC to each humidifying well in the Hybridization Chamber
3. Place the slide in a clean hybridization chamber.
4. Add cover slip ("lifter slip") on top of the array.
5. Add pre-hybridization solution until the entire area under the cover-slip is covered.
6. Put on the top of the hybridization chamber and attach margin clamps.
7. Immerse the chamber into a 42°C water bath and pre-hybridize 1 hour.
8. Wash slide with H_2O 1 min.
9. Wash slide with 100% isopropanol 1min.
10. Spin dry slides in centrifuge 5 min at 500 rpm

C. Hybridisation:

1. Adjust the probe volume to 34 µl with ddH_2O
2. Add 9.2 µl 20XSSC
   5.2 µl 2.5% SDS
   3.3 µl tRNA
   3.3 µl Cot-1DNA
   3.9 µl A_{80} oligo 0.85µg/µl
3. Denature probe by heating for 2 min at 100°C.
4. Incubate at 37°C for 10 min.
5. Centrifuge the samples 10s at maximum speed.
6. Place the pre-hybridized slide in a clean hybridization chamber.
7. Add cover slip(“lifter slip”) on top of the array.
8. Carefully pipette the probe to the edge of the cover-slip.
9. Pipette a drop of 15 µl 3XSSC on the slide.
10. Put on the top of the hybridization chamber.
11. Immerse the chamber into a 65°C water bath and hybridize for 14 -18 hours.
12. Remove chamber from bath and array from chamber.
13. Carefully place array into slide holder in washing solution 1 at 55°C.
14. Shake array slowly until cover slip falls off 5 min.
15. Plunge slide ≈20 times (5 min) and transfer the slides one by one to a second chamber with washing
    solution 2 at 55°C for 5 min. Plunge ≈20 times and transfer the slides to a third chamber with washing solution 3 at 55°C for 5 min, plunge ≈20 times.
16. Spin array dry in centrifuge at 500 rpm for 5 min
17. Scan array