Sampling Effects on Gene Expression Data
from a Human Tumour Xenograft

by *Berner JM[1], Müller CR[1], Holden M[2], Wang J[1], Hovig E[1] and Myklebost O[1,*]*

[1] Department of Tumour Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, Norway.

[2] Norwegian Computing Center, Oslo, Norway.

Summary

Human tumour tissue transplanted to and passed through immunodeficient mice as xenografts make powerful model systems to study tumour biology, in particular to investigate the dynamics of treatment responses, e.g. to chemotherapeutic agents. Before embarking on large-scale gene expression analysis of chemotherapy response in human sarcoma xenografts, we investigated the reproducibility of expression patterns derived from such samples. We compared expression profiles from tumours from the same or different mice and of various sizes, as well as central and peripheral parts of the same tumours. Twenty-three microarray hybridisations were performed on cDNA arrays representing 13000 genes, using direct labelling of target cDNAs. An ANOVA-based linear mixed-effects model was constructed, and variances of experimental and biological factors contributing to variability were estimated. With our labelling procedure used, the effect of switching the dyes was pronounced compared to all other factors. We detected a small variation in gene expression between two tumours in the same mouse as well as between tumours from different mice. Furthermore, central or peripheral position in the tumour had only moderate influence on the variability of the expression profiles. The biological variability was comparable to experimental variability caused by labelling, confirming the importance of both biological and technical replicates. We further analysed the data by pair-wise Fisher's linear discriminant method and identified genes that were significantly differentially expressed between samples taken from peripheral or central parts of the tumours. Finally, we evaluated the result of pooling biological samples to estimate the recommended number of arrays and hybridisations for microarray experiments in this model.

Introduction

The passing of human tumour tissue through a series of immunodeficient nude mice ("serial passaging") has been an important model system in cancer research for more than 40 years (*Rygaard & Povlsen, 1969*). Such xenograft models have the advantage of unlimited availability of live tumour tissue, allowing for repeated experiments on the same tumour. It has been shown by various techniques, including histology and DNA ploidy, that the transplants are stable over many passages and for several years (*Povlsen et al., 1980*). For most practical purposes it has been assumed that multiple xenografts, derived from a single xenograft sample, are more or less identical within a limited number of passages. Limitations of the system may be due to selection bias of tumour lines by their transplantability, the replacement of the human stroma by murine stroma cells and altered host/graft interaction as for example regarding drug metabolism and immunology. These limitations have to be addressed when interpreting the results of xenograft experiments.

At our institution xenograft models are used in a

Correspondence: Ola Myklebost, Department of Tumour Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Montebello, 0310 Oslo, Norway
Tel: +47 22 93 42 99
Fax: +47 22 93 24 21
E-mail: olam@radium.uio.no

broad range of settings, more recently including cDNA microarray analysis for both tumour profiling and monitoring the response to chemotherapeutic agents or other experimental treatments. The approach assumes homogeneity of multiple transplants originating from the same parent tumour. Our primary aim with this study was to estimate the inherent biological variability of the xenograft model system and to compare it to the variability in the microarray technique. This knowledge will be used to design experiments that provide high quality reproducible data by simultaneously decreasing the costs and reducing the number of sacrificed mice. We were particularly interested in the variation between transplants in the same mouse, or in different mice, of transplants of varying size, and of central and peripheral parts of the same tumour. Vascularisation and growth rate is thought to be higher in the tumour periphery than in central parts, and might differ between larger and smaller tumours. We wanted to determine to what extent expression patterns would vary with position within the tumour, as this could affect the sampling technique and limit the tissue harvest. Furthermore, we estimated the minimum number of arrays and hybridisations required for experiments in our mouse model, and evaluated whether it is advantageous to pool several tumours from different mice instead of replicating the experiment.

A common way to estimate uncertainties caused by different sources of variation, is to fit an Analysis of Variance (ANOVA) -based statistical model to the data. A typical model in the microarray setting would include variables such as dye, array, gene and appropriate interaction of these terms. Further variables would be included according to the experimental setting, in our case variation caused by tumour, mouse, size and position in tumour, and interaction between these and the former variables.

Material and Methods
Animals
All procedures involving animals were performed according to protocols approved by the Animal Care and Use Committee at the hospital and in compliance with the National Ethics Committee's guidelines on animal welfare and the Council Directive of the European Communities on the protection of animals used for experimental and other scientific purposes.

Female nude mice (Balb/c: nu/nu, NCI) were bred in our animal facility, weaned after 21 days and maintained in a specific pathogen-free environment at controlled temperature (21 +/- 0.5°C) and humidity (55-65 %) on an unreversed 12 hour light cycle (light 0700-1900). Sentinels were tested according to FELASA's health monitoring recommendations. Groups of maximal eight mice were kept in transparent polypropylene cages (Tecniplast Eurostandart type III, Scanbur BK, Nittedal, Norway) on aspen chips bedding (B&K Universal, Hull, UK) with pelleted food (RM3, Special Diets Services, Witham, UK) and acidified water supplied *ad libitum*. Morbidity was controlled for by daily inspections focusing on behaviour, posture or weight loss; >10 % weightloss was a humane endpoint. The animals were anesthetized by intraperitoneal injection of 0.05-0.1ml of a solution containing 0.08 mg/ml fentanyl, 2.5 mg/ml fluanison (Janssen Pharmaceutica, Beerse, Belgium) and 1.25 mg/ml midazolam (Roche, Basel, Switzerland) during transplantation and killed by cervical dislocation before tumour harvest.

Tissue samples
A human leiomyosarcoma xenograft (LMS2x) was established from a grade three intramuscular tumour on the forearm by direct subcutaneous transplantation to nude mice, five to seven weeks of age, and maintained by serial passaging; latency was 21 days. Histologically, the xenograft resembles a spindle cell sarcoma consistent with a high grade malignant leiomyosarcoma (Figure 1). A total of nine tumour samples from six mice were analysed (Table 1). A tumour was considered as small if the volume was less than 750 mm^3 (calculated by the formula 0.5 x length x width^2) and large if the volume was more than 750 mm^3. From mouse num-

ber 3, two tumours (T3-I and T3-II) were available, one from each flank of the mouse. Tumours T4 and T5, from mice 4 and 5, were dissected into both peripheral and central parts, and from tumour T6 only the peripheral parts were analysed. The tissues were snap-frozen in liquid nitrogen immediately after harvesting.

quantified by spectrophotometry and examined by 1 % agarose gel electrophoresis. Intact RNA was then extracted twice in 1 ml TRIzol reagent (Life technologies, Grand Island, NY). Xenograft tissue from three different passages from the same sample was combined and applied as a reference for the microarray analysis.
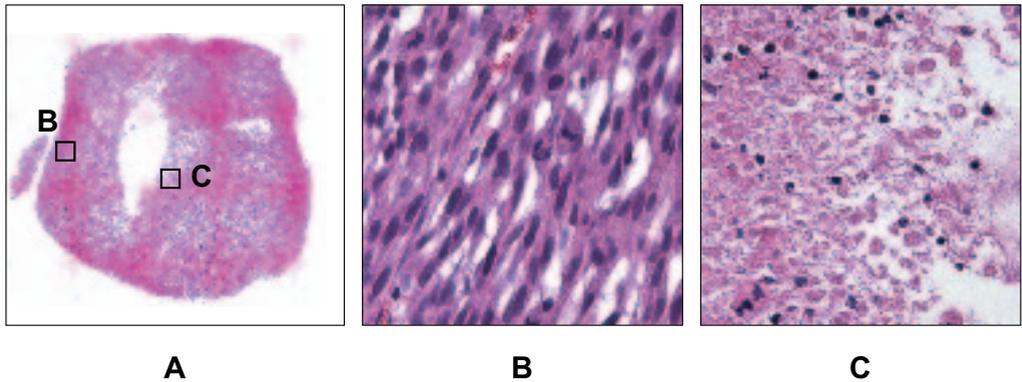

Figure 1. Histological section of LMS2x.
A) Whole tumour section; B) Vital tumour periphery; C) Necrotic tumour centre.

Table 1.
Samples from xenograft LMS2x

Tumour	Mouse	Tumour volume	Size tumour	Samples	Part of tumour
T1	1	446	small	T1	whole
T2	2	253	small	T2	whole
T3-I	3	600	small	T3-I	whole
T3-II	3	488	small	T3-II	whole
T4	4	2601	large	T4-per	peripheral
				T4-cen	central
T5	5	850	large	T5-per	peripheral
				T5-cen	central
T6	6	926	large	T6-per	peripheral

Tumour volume in mm^3

RNA preparation and labelling
Total RNA was isolated using Qiagen RNeasy Midi kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instruction. Extracted RNA was

Cy3 and Cy5 labelled cDNA probes were prepared using the CyScribe First Strand cDNA Labelling kit (Amersham Pharmacia Biotech AB, Uppsala, Sweden), mainly according to the manufacturer's

instruction. Briefly, 30 μg of total RNA was mixed with 1 μl anchored oligo-dT and 1 μl random nonamer primers, heated to 70 °C for 5 minutes and then cooled to room temperature for 10 minutes to allow the primers to anneal with the template. Reverse transcription with incorporation of Cy3-dUTP and Cy5-dUTP (Amersham Pharmacia) was done as described in the manual. The reaction mixture was incubated at 42 °C for 1.5 hours and then stopped by adding 2.5 μl 0,5 M EDTA (pH = 8.0). Residual RNA was hydrolysed by adding 5 μl 1 M NaOH and incubated for 30 minutes at 65 °C. Ten μl of 2 M HEPES was added to neutralize the reaction mixture. Micro Bio-Spin P-6 columns (Bio-Rad Laboratories, Hercules, CA) were used to remove unincorporated nucleotides. The Cy3- and Cy5-labelled cDNAs were combined, 15 μg human Cot-1 DNA (Gibco BRL) was added, and the sample was concentrated using Microcon 30 columns (Amicon, Millipore Corporation, Bedford, MA).

Hybridisation
The cDNA microarrays were produced in house using a Micro Grid II robotic printer (Bio Robotics, Cambridge, UK). Human cDNA arrays containing thirteen thousand (13k) probes were printed on amino-silane coated slides (CMT GAPS II, Corning Life Sciences, Corning, NY), see http://www.mikromatrise.no. All the slides used were from the same print series. The hybridisation solution (total 40 μl) consisted of the Cy3 and Cy5-labeled cDNA, 3.5 x SSC (pH=7.5), 0.3 % SDS, 1.25 x Denhardt's solution, 8 μg yeast tRNA and 16 μg poly dA (Amersham Pharmacia Biotech AB). The final mixture was heated for 3 min at 100 °C and spun down for 10 min at 12 000 g before it was applied to a microarray slide under a Lifter Slip cover glass (Erie Scientific Company, Portsmouth, NH). The slide was then placed in an ArrayIT hybridization chamber (Telechem, Sunnyvale, CA) and submersed in a water bath for overnight hybridization at 65 °C. The coverslip was removed in a solution of 2 x SSC and 0.1% SDS. The slide was washed at room temperature in 1 x SSC for 5 minutes,

0.2 x SSC for 5 minutes and 0.05 X SSC for 2 minutes, and then dried by centrifugation.

Imaging and data analysis
The hybridised slides were scanned with a ScanARRAY 4000 (Packard Biosciences, Biochip Technologies LLC, Meriden, CT), and the obtained signal intensities were quantified with GenePix Pro 3.0 software (Axon Instruments Inc., Union City, CA). A background-corrected intensity for each spot was calculated by subtracting the median of the pixels in the local background from the mean of the pixels in the spot. We manually removed technically flawed spots and spots flagged by GenePix. Finally, all spots where the spot intensity (uncorrected foreground intensity) was smaller than the median of the background intensity plus two times background standard deviation were removed (i.e. spot intensity < median [background intensity + 2x(standard deviation background)], in any of the two channels (*Wang et al., 2002*). An overview of the number of filtered genes is presented in Table 2. The data were normalized using locally weighted scatterplot smoothing (LOWESS), as described in Yang et al. (*Yang et al., 2002*), and implemented in the sma package in the statistical language R. This routine normalized the ratios for each spot on each array by subtracting a value s, depending on the value of the log2-transformed total intensity of the spot on the array. To identify genes that were significantly differentially expressed between samples taken from peripheral or central part of the tumours, we used the pair-wise Fisher's linear discriminant (PFLD) method (*Johnson & Wichern, 1998; Wang et al., 2003*).

ANOVA-based statistical modelling and experimental design
ANOVA-based statistical modelling is a means of evaluating the source of variability and their interactions. We identified the following eight main effects or factors for the design: The RNA samples were labelled red (Cy5) or green (Cy3), so the dye effect (D) had two levels, one level when the computed

Table 2.
Filtering per array

Sample	Sample labeled with dye	Genes flagged by GenePix	Genes manually flagged	Additional filtering[1]	Total number of removed genes	Number of genes after filtering[2]
T1	Cy3	4839	26	3635	8500	4940
T1	Cy3	3018	29	2937	5984	7456
T2	Cy3	5741	12	3916	9669	3771
T2	Cy3	1578	42	1754	3374	10066
T3-I	Cy3	6952	27	4368	11347	2093
T3-I	Cy3	3279	39	3892	7210	6230
T3-II	Cy3	1745	36	1725	3506	9934
T3-II	Cy3	2281	47	3085	5413	8027
T4-per	Cy3	3190	1	1316	4507	8933
T4-per	Cy3	2026	7	1443	3476	9964
T4-cen	Cy3	3046	32	3400	6478	6962
T4-cen	Cy3	2690	30	3687	6407	7033
T5-per	Cy3	5620	30	3534	9184	4256
T5-per	Cy3	3040	23	3462	6525	6915
T5-cen	Cy3	2556	16	2804	5376	8064
T5-cen	Cy3	2844	18	3248	6110	7330
T6-per	Cy3	3451	8	4327	7786	5654
T6-per	Cy3	6764	12	3552	10328	3112
T6-per	Cy3	4460	4	3654	8118	5322
T6-per	Cy3	5185	12	3586	8783	4657
T6-per	Cy5	4050	17	4441	8508	4932
T6-per	Cy5	1916	55	2485	4456	8984
T6-per	Cy5	1169	23	880	2072	11368

[1] Minimum intensities for Cy3 and Cy5 channels: median [background intensity + 2 x (standard deviation background)]

[2] Total number of genes: 13440

ratio was red/green and one when the ratio was green/red. The ratio is always the intensity for a tumour sample divided by the intensity for the reference. We analysed material from six different mice (M) and seven different tumours (T). Tumours were classified as either large or small, i.e. the size (S) effect had two levels. Material was taken from the centre or the periphery of the tumour, i.e. the position (P) effect had two levels as tissues from small tumours were classified as being taken from the periphery. Twenty-three different hybridisations (H) were performed. RNA was labelled separately for each hybridisation, except for one green and one red labelling, which were used for three hybridisations (i.e. three labelling reactions were mixed). This means that the labelling (L) effect had 21 levels. A total of 11140 genes (G) remained after filtering (each of these genes was present on at least one microarray). For an overview of the experimental design, see Figure 2.
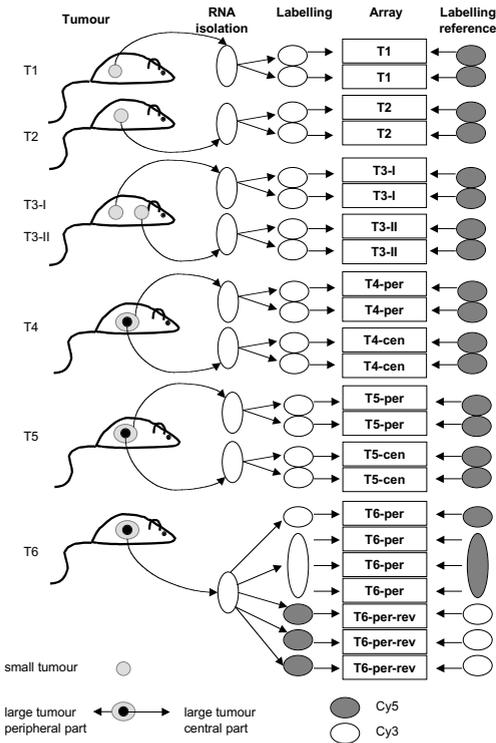
Figure 2. Experimental design of the 23 hybridisations

Two separate microarray assays were performed for each sample. The samples were labelled with Cy3-dUTP and the reference with Cy5-dUTP. One sample (T6-per) is represented by seven arrays. The labelled cDNA probes were mixed before applying to three slides (T6-per). For three other hybridisations (T6-per-rev), the labelling scheme was reversed to control for dye-based bias.

The ANOVA-based linear mixed-effects model

To investigate the different sources of variability, we set up an ANOVA-based statistical model. (Related models are found in *Kerr et al., 2000, Wolfinger et al., 2001, Jin et al., 2001* and in *Nygaard et al., 2003*).

Let $y_{mtsplhdg}$ denote the log$_2$-ratio for gene g in hybridisation h. The material used is from labeling l, for which dye Cy5 and Cy3 have been chosen accor-

ding to the value of d. The mouse material was taken from position p in tumour t in mouse m. This tumour has size s. We explain the calculated log$_2$-ratio by the following linear mixed-effects model:

$$y_{mtsplhdg} = \mu + (-1)^d D + (-1)^s S + (-1)^p P + M_m + T_t + L_l + H_h + G_g + (-1)^d DG_g + (-1)^s SG_g + (-1)^p PG_g + MG_{mg} + TG_{tg} + LG_{lg} + HG_{hg} + \varepsilon_{mtsplhg}$$

where $m=1,...,6$, $t=1,...,7$, $s = 1$ (small) or 2 (large), $p = 1$ (peripheral) or 2 (central), $l = 1,…,21$, $h=1,…,23$, $d=1$ if the ratio is red/green, and 2 if green/red, and $g=1,…,11140$. μ is the overall mean. M_m, T_t, L_l, and H_h are the overall effects of mouse m, tumour t, labelling l, and hybridisation h, respectively. The overall dye effect of dye d is $-D$ when $d=1$, and D when $d=2$. Similarly, for the size effect S and s, and for the position effect P and p. Furthermore, MG_{mg}, TG_{tg}, LG_{lg} and HG_{hg}, are the gene-specific mouse, tumour, labelling and hybridisation effects, respectively. The gene-specific dye effect is $-DG_g$ when $d=1$, and DG_g when $d=2$. Similarly, for the gene-specific size effect SG_g and s, and for the gene-specific position effect PG_g and p.

Only μ, D, S and P were modelled as fixed effects. All others effects and interaction effects were modelled as independent normally distributed random effects with zero mean and constant standard deviation. This means that $M_m \sim N(0,\sigma_M^2)$, $T_t \sim N(0,\sigma_T^2)$, and similar for all the other random effects. Also the modelling and measurement error is assumed to be normally distributed with zero mean and constant variance, i.e. $\varepsilon_{mtsplhg} \sim N(0,\sigma_\varepsilon^2)$. Due to the large dimensions in the mixed-effects model, we estimated the parameters of interest (the fixed effects and the variances of the random effects) using Gibbs sampling. Gibbs sampling is an iterative simulation method, which is a particular algorithm in a wider class of such algorithms, named Markov chain Monte Carlo methods (*Gilks et al., 1996; Robert & Casella, 1999*). Follestad et al. (*2004*) give further details and a description of the C-program used for estimating the model parameters (the C-program is available from Follestad et al. (*2004*), on request).

Having estimated the standard deviations of the random effects, these estimates could be used for computing the uncertainty of the ratios obtained from one hybridisation using material from one tumour. To obtain ratios with less uncertainty we could take the average of calculated log$_2$-ratios for the same spot on different arrays.

Experimental design and effect of pooling
Another way of reducing gene expression uncertainty in xenograft experiments and minimize the effect of biological variation, is to pool RNA from several tumours into one sample. Based on the present experimental data and the defined model, we have evaluated whether it is advantageous to reduce the number of experiments by pooling RNA from several xenografts. We will only consider experiments where a reference is used, where the reference is always labelled with the same dye, and where material is always taken from the periphery of the tumour. Consequently, the standard deviations concerning dye and position are left out from the formulas given below. We also leave out the standard deviations concerning size because it turned out that size is not an important effect (see the Results section). In the experimental designs sketched below, let n_M be the number of mice contributing to each biological sample. Let n_H be the number of hybridisations in the design and n_S the number of biological samples used. We allow $n_S = 1$ or n_H, i.e. either a common sample is used for all arrays or one sample is used for each hybridisation. Let n_L be the number of labelling with each dye in the design. We allow $n_L = n_H$ or n_S, i.e. either one labelling of each sample or one labelling for each hybridisation. The log$_2$-ratio for each spot is found by averaging the n_H log$_2$-ratios for this spot. The standard deviation for biological variation, after pooling of material from several mice, σ^2_{sample,n_M}, is then found by the following simulation:
Compute the square root of the sample variance of, $b_1,...,b_{100000}$ where

$$b_k = \log_2\left(\frac{\sum_{i=1}^{n_M} 2^{m_{i,k} + mg_{i,k} + t_{i,k} + tg_{i,k}}}{n_M}\right)$$

and where $m_{i,k}$, $mg_{i,k}$, $t_{i,k}$ and $tg_{i,k}$, are random numbers drawn from $N(0,\sigma^2_M)$, $N(0,\sigma^2_{MG})$, $N(0,\sigma^2_T)$ and $N(0,\sigma^2_{TG})$, respectively. The standard deviation for the log$_2$-ratios obtained by averaging replicates will then be

$$\sigma^2_{total,n_M,n_S,n_L,n_H} = \frac{\sigma^2_{sample,n_M}}{n_S} + \frac{\hat{\sigma}^2_L}{n_L} + \frac{\hat{\sigma}^2_H}{n_H} + \frac{\hat{\sigma}^2_{LG}}{n_L} + \frac{\hat{\sigma}^2_{HG}}{n_H} + \frac{\hat{\sigma}^2_\varepsilon}{n_H}$$

Results and Discussion
Experimental outline
The main objective of the current study was to investigate the biological and experimental variability of the expression profiles of xenograft transplants. Global gene expression was examined in nine samples from seven tumours originating from the same leiomyosarcoma xenograft as outlined in Table 1. Figure 2 presents an overview of the experimental design for the 23 hybridisations. Two separate microarray assays were performed for each sample where the test sample was labelled with Cy3-dUTP and the reference with Cy5-dUTP. One sample (T6-per) is represented by seven arrays. For three hybridisations, the labelling scheme was reversed to control for dye-based bias. For three other slides, the labelled cDNA probes were mixed before applied on three separate slides, to be able to distinguish between the gene expression variation introduced by the process of either labelling or hybridisation. Data from the 23 cDNA microarray hybridisations were analysed as described in Material and Methods.

The ANOVA-based linear mixed-effects model
An ANOVA-based model was constructed to investigate the sources of variation and how they contribute to changes in gene expression values. We evaluated the variations in expression profiles between transplants from the same or different mice, of varying size or from central and peripheral parts of

the same tumour. Furthermore, we analysed experimental variables influencing gene expression like dye (Cy3 or Cy5), the labelling of the probe and the hybridisation. The estimates for the standard deviations of the random effects in the model are presented in Table 3. The standard deviation of mouse, tumour, labelling and hybridisation effects are (as expected) small, because the data have been normalised. The test and the reference in each hybridisation were samples from the same xenograft, and consequently the gene effect σ_G is low too (0.013). The modelling and measurement error, σ_E, was moderate.

The estimated variability in gene expression caused by the labelling (σ_{LG}) is moderate (0.123), indicating that labelling efficiency in the reverse transcription reaction and further handling of the probe contributes to a small level of variability. This emphasizes the importance of performing replicate experiments (see below for calculation of number of replicates needed). The estimated gene-specific variability caused by hybridisation (σ_{HG}) is low (0.013), showing that the process of hybridisation itself or differences of the arrays do not contribute

much to variability. σ_{LG} is considerably more important than σ_{HG}, demonstrating that gene expression varies more with labelling than with array and hybridisation. However, others have shown that even the spot itself is subject to substantial variability and have emphasized the need for experimental replication to achieve reliable gene expression values (*Lee et al., 2000*).

The most prominent factor contributing to variation in gene expression is switching of dyes (σ_{DG} = 0.423). It is known that polymerases introduce modified nucleotides with different efficiency dependent on the sequence context, and different enzymes are more or less affected by the steric effects of the Cy5 or Cy3 labelled nucleotides. For the generation of Cy3 and Cy5 labelled cDNA probes, we used the CyScript reverse transcriptase. In a related study performing direct labelling with superscript reverse transcriptase, the estimated standard deviation of the gene and dye interaction was small (*Nygaard et al., 2003*). Indirect labelling methods, where e.g. biotin is used to label the nucleotides and the fluorochromes are bound to biotins

Table 3.

Results for the linear mixed-effects model. Estimated standard deviations with 95% symmetric credibility intervals (C.I.)

Parameter	Description	Estimate (95% C.I.)
σ_M	Mouse	0.047 (0.019, 0.109)
σ_T	Tumour	0.050 (0.021, 0.109)
σ_L	Labelling	0.033 (0.018, 0.056)
σ_H	Hybridisation	0.035 (0.018, 0.060)
σ_G	Gene	0.013 (0.010, 0.017)
σ_{MG}	Gene expression varies with mouse	0.147 (0.145, 0.150)
σ_{TG}	Gene expression varies with different tumours from the same mouse	0.128 (0.125, 0.131)
σ_{SG}	Gene expression varies with size of tumour	0.018 (0.013, 0.024)
σ_{PG}	Gene expression varies with position in tumour	0.092 (0.089, 0.095)
σ_{LG}	Gene expression varies with labelling	0.123 (0.120, 0.126)
σ_{HG}	Gene expression varies with hybridisation and array	0.013 (0.010, 0.018)
σ_{DG}	Gene expression varies with dye	0.423 (0.419, 0.428)
σ_ε	Error	0.130 (0.128, 0.132)

after enzymatic incorporation, can be applied to overcome this problem.

Few studies have investigated variations in global gene expression between multiple transplants from the same cell-line. For the xenograft LMS2x, our results show that the estimated standard deviation of the interactions between gene and mouse σ_{MG} (0.147) and gene and tumour σ_{TG}, (0.128) are modest. This indicates that a low level of variation in expression profiles can be expected between tumours originating from this xenograft, even if they are derived from the same mouse. The observed biological differences between the xenografts are small, but still imply that biological replicates are essential to obtain convincing data (*Chen et al., 2004; Yang & Speed, 2002*), preferably using samples from different animals.

The gene expression did not vary significantly with size of the tumour (σ_{SG} = 0.018). However, the analysis indicates that the position in tumour, either centre or peripheral, has a moderate influence on gene expression variability (σ_{PG} = 0.092). Xenografts are known to display heterogeneous vascular densities including avascular areas (*Konerding et al., 1999*), and an inverse correlation between tumour size and necrosis has been reported (*Harrington et al., 2000*). We observed necrosis in all but very small tumours (<250 mm³), covering a substantial part of larger tumours, more predominant in the tumour centre but also in its periphery (Figure 1). Thus, the histological findings support the results obtained for the linear mixed-effects model, and indicate that it is not essential for gene expression studies in the present xenograft-model, to select tumours of identical size or from a specific part of the tumour.

Differentially expressed genes in peripheral and central part of xenografts
Even if the differences in expression profiles are small between the outer and inner parts of these xenografts, we expected to find a limited number of genes with altered expression within samples taken from peripheral or central parts of the tumours. We

thus searched for genes that could discriminate between these parts of the tumours using the pairwise Fisher's linear discriminant (PFLD) method and found 60 genes to be differentially expressed. Here, we present the log2-ratios of the 15 most significant genes (Figure 3). One of the most significantly differentially expressed genes in our PFDL analysis was *COL3A1*, coding for type III procollagen alpha. Collagens are extracellular structural protein involved in formation of connective tissue, and it has been suggested that type III collagen is important for the development of skin and the cardiovascular system (*Liu et al., 1997*). We found higher expression of *COL3A1* in the peripheral sections compared with central, and we can only speculate if this difference is caused by less vascular activity in central parts. In another gene expression study, *COL3A1* was the most upregulated gene in advanced carcinomas compared to local carcinomas (*Tapper et al., 2001*). Hypoxia has been suggested to result in a more metastatic phenotype (*Brizel et al., 1996*), and this observation could be related to this phenomenon. Increased expression of fibroblast growth factor 2 (*FGF2*) was found in peripheral parts compared to central part. Interestingly, it was demonstrated that hypoxic endothelial cells had an increased responsiveness to FGF2 due to a HIF-1α-dependent increase in FGF2-binding heparan sulphate sequences (*Li et al., 2002*). Gamma-aminobutyric acid A receptor, gamma 2 (GABRG2) was also found to be overexpressed in peripheral parts. GABA$_A$ receptors mediate synaptic inhibition in the mammalian brain and have been detected on non-neuronal cells including the uterus (*Hedblom & Kirkness, 1997*) and breast cancer (*Jiang et al., 2002*), but their role in the latter systems remain unclear.

As seen in Figure 3, even genes selected for significant expression displayed only minor change, indicating that the various parts of these xenografts are quite similar with regard to gene expression. If we assume that the central part of the tumour is more oxygen-deprived, hypoxia will alter gene
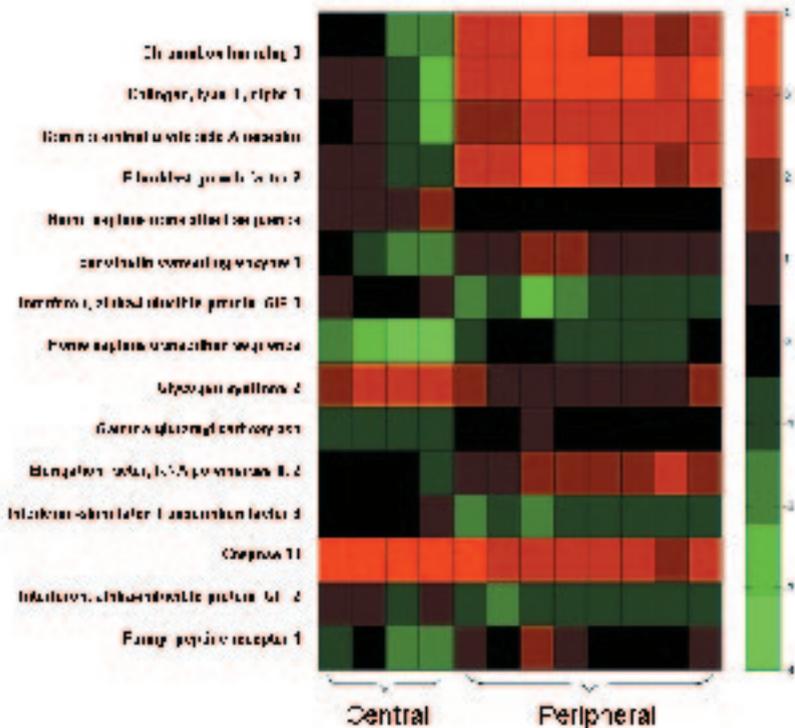
Figure 3. Overview of genes discriminating between the peripheral and central parts of the tumours
A presentation of the 15 most differentially expressed genes from the central and peripheral parts of the xenografts is given. The genes are listed with decreasing score. The log2-ratio intensities are as indicated to the left, red colour boxes represent genes with higher expression in the sample than in the reference and green boxes represent reduced expression compared to the reference.

expression to improve blood supply, to adapt metabolism, and protect against cellular damage. Several pathways are known to be regulated by hypoxia, including angiogenesis, glycolysis, growth factor signalling, immortalisation, genetic instability, tissue invasion and metastasis, apoptosis and pH regulation. The key transcription factor affecting gene regulation is hypoxia-inducible factor-1α (HIF-1α) (*Wang et al., 1995*), which can activate several genes, including vascular endothelial growth factor (*VEGF*), which promotes angiogenesis; solute carrier family 2 (facilitated glucose transporter member 1) (*SLC2A1*), which activates glucose transport; lactate dehydrogenase (*LDHA*),

which is involved in the glycolytic pathway; and erythropoietin (*EPO*), which induces erythropoiesis (*Harris, 2002*). None of these most known hypoxia-induced genes was identified in our PFLD analysis. However, considerable variation in the gene expression levels of both *VEGF* and *SLC2A1* was found between the different xenograft samples. Interestingly, the lowest level of expression of these genes was found in T4-per and T5-per, two peripheral samples, and T4-cen and T5-cen showed higher levels (Figure 4). The gene expression of both *HIF-1α* and *LDHA* was stable between the various samples, and *EPO* was not present on the arrays.
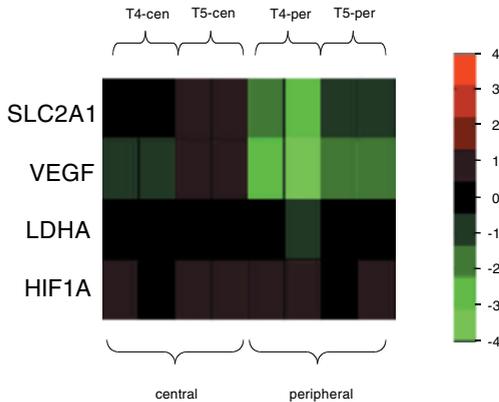
Figure 4. Gene expression of specific genes involved in hypoxia between peripheral and central part of tumour.

The log2-ratio intensities are as indicated to the left, red colour boxes represent genes with higher expression in the sample than in the reference and green boxes represent reduced expression compared to the reference.

In a previous report, a few differentially expressed genes were found in central and peripheral parts of a leiomyosarcoma patient sample (*Shmulevich et al., 2002*). However, none of these differentially expressed genes showed significant changes in our samples. Minimal intratumour variations were found with gene expression profiling. In another study, core biopsies from a Ewing sarcoma xenograft and a neuroblastoma xenograft and corresponding excision biopsies were shown to have conserved expression profiles (*Sotiriou et al., 2002*). These authors suggested that even if small differences between individual genes exist, the composite expression of the core biopsies is representative of the entire tumour.

Experimental design and effect of pooling
One approach to reduce gene expression uncertainty in xenograft experiments and to minimize the effect of biological variation is to pool RNA from several tumours into one test sample. Studies have revealed that pooling of biological samples is statistically valid, and that appropriate RNA pooling can provide equivalent power and reduce the costs of microarray experiments (*Kendziorski et al., 2003; Peng et al., 2003*). It has been suggested that pooling is particularly advantageous if the biological variation is high compared to the technical variation on the arrays (*Kendziorski et al., 2003*). Based on the present experimental data and the defined model, we have evaluated if it is advantageous to reduce the number of experiments by pooling RNA from several xenografts (see Material and Methods for more details). Here we consider some selected designs, comparing the effect of pooling various numbers of tumours and performing replicate hybridisations (Figure 5). The estimated standard deviation of the $\log_2$-ratio for any gene analysing one tumour on one array is 0.277. By examining three tumours by three arrays without pooling, the standard deviation is considerably reduced (0.160). In comparison, by pooling the three samples and performing three hybridisations, standard deviations are identical (0.160). Performing six hybridisations generates a standard deviation at a similar level as without pooling using three tumours and duplicate experiments (six arrays). However, if five tumours are pooled and six hybridisations performed, the standard deviation is reduced to 0.120.

In general, to obtain a low standard deviation and simultaneously reduce the number of arrays, the number of tumours included in the pool should be increased. For example, if eight tumours are pooled, three hybridisations will give an estimated standard deviation of 0.130 and a better reproducibility compared to three tumours hybridised separately but in duplicate (0.141). In most cases, this is not a realistic design, as too many animals have to be sacrificed. However, this analysis shows that if many tumours are available, they should be included and pooled.
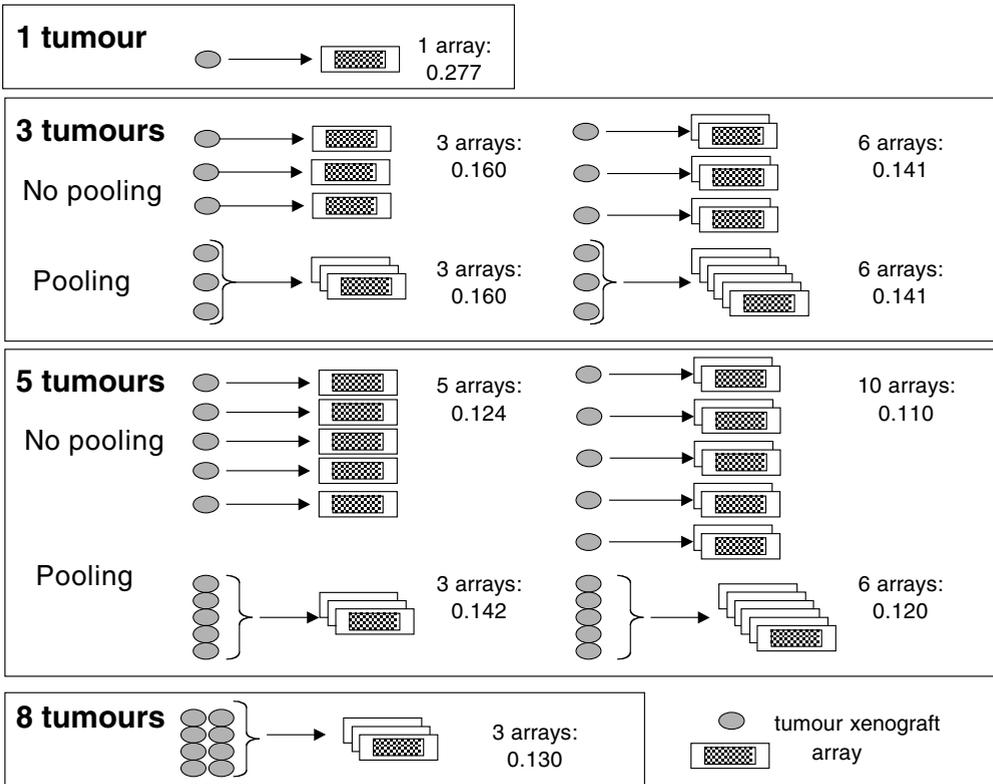
Figure 5. Effect of pooling biological samples
The figure presents the estimated standard deviations for the log2-ratios in various experimental designs. The effect of pooling, increasing the number of biological samples or including more arrays is illustrated.

Conclusions
The statistical analysis revealed that both experimental and biological factors represent a considerable cause of variation in microarray hybridisation experiments, confirming the need for both biological and technical replicates. The most prominent parameter contributing to variability in the expression profiles was the dye, showing that our labelling method has a profound effect on the results. Selecting tumours of identical size or samples with a specific tumour position is not essential when performing gene expression studies in the xenograft model LMS2x. However, when interpreting expression array data in the model, it should be conside-red that a small number of genes are affected by tumour heterogeneity. Pooling of biological samples can be favourable to counteract this phenomenon.

References

Brizel DM, SP Scully, JM Harrelson, LJ Layfield, JM Bean, LR Prosnitz and MW Dewhirst: Tumor oxygenation predicts for the likelihood of distant metastases in human soft tissue sarcoma. Cancer Res 1996, *56*, 941-943.

Chen J , R Delongchamp, C Tsai, H Hsueh, F Sistare, KL Thompson, VG Desai and JC Fuscoe: Analysis of variance components in gene expression data. Bioinformatics 2004, *20*, 1436-1446.

Follestad T, M Langaas, H Rue, M Holden and A Løland: glme: a C-program for parameter estimation using Gibbs-sampling in large linear-mixed-effect models, with applications to DNA microarray data. NR note SAMBA/10/04, *2004*. http://publications.nr.no/S2-2004.pdf

Gilks WR, S Richardson and DJ Spiegelhalter: Marchov Chain Monte Carlo in Practice. Chapman & Hall, London, 1996.

Harrington KJ, G Rowlinson-Busza, KN Syrigos, RM Abra, PS Uster, AM Peters and JS Stewart: Influence of tumour size on uptake of (111)ln-DTPA-labelled pegylated liposomes in a human tumour xenograft model. Br J Cancer 2000, *83*, 684-688.

Harris AL: Hypoxia - a key regulatory factor in tumour growth. Nat Rev Cancer 2002, *2*, 38-47.

Hedblom E and EF Kirkness: A novel class of GABAA receptor subunit in tissues of the reproductive system. J Biol Chem 1997, *272*, 15346-15350.

Jiang Y, SL Harlocker, DA Molesh, DC Dillon, JA Stolk, RL Houghton, EA Repasky, R Badaro, SG Reed and J Xu: Discovery of differentially expressed genes in human breast cancer using subtracted cDNA libraries and cDNA microarrays. Oncogene 2002, *21*, 2270-82.

Jin W, RM Riley, RD Wolfinger, KP White, G Passador-Gurgel and G Gibson: The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster. Nat Genet 2001, *29*, 389-395.

Johnson R and D Wichern: Applied multivariate statistical analysis. Prentice-Hall, New Jersey; 1998.

Kendziorski CM, Y Zhang, H Lan and AD Attie: The efficiency of pooling mRNA in microarray experiments. Biostatistics 2003, *4*, 465-477.

Kerr MK, M Martin and GA Churchill: Analysis of variance for gene expression microarray data. J Comput Biol 2000, 7, 819-837.

Konerding MA , W Malkusch, B Klapthor, C van Ackern, E Fait, SA Hill, C Parkins, DJ Chaplin, M Presta and J Denekamp: Evidence for characteristic vascular patterns in solid tumours: quantitative studies using corrosion casts. Br J Cancer 1999, *80*, 724-732.

Lee ML, FC Kuo, GA Whitmore and J Sklar: Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci USA 2000, *97*, 9834-9839.

Li J, NW Shworak and M Simons: Increased responsiveness of hypoxic endothelial cells to FGF2 is mediated by HIF-1alpha-dependent regulation of enzymes involved in synthesis of heparan sulfate FGF2-binding sites. J Cell Sci 2002, *115*, 1951-1959.

Liu X, H Wu, M Byrne, S Krane and R Jaenisch: Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. Proc Natl Acad Sci USA 1997, *94*, 1852-1856.

Nygaard V, A Løland, M Holden, M Langaas, H Rue, F Liu, O Myklebost, Ø Fodstad, E Hovig and B Smith-Sørensen: Effects of mRNA amplification on gene expression ratios in cDNA experiments estimated by analysis of variance. BMC Genomics 2003, *4*, 1471-2164.

Peng X, C Wood, E Blalock, K Chen, PW Landfield and A Stromberg: Statistical implications of pooling RNA samples for microarray experiments. BMC Bioinformatics 2003, *4*, 26.

Povlsen CO, M Spang-Thomsen, J Rygaard and J Visfeldt: Heterotransplantation of human malignant tumours to athymic nude mice   In: Immunodeficient animals for cancer research,

ed: S Sparrow, MRC Laboratory Animals Centre, Carshalton, Surrey, 1980, 95-103.

Robert CP, and G Casella: Monte Carlo statistical methods. Springer-Verlag, New York; 1999.

Rygaard, J and CO Povlsen: Heterotransplantation of a human malignant tumour to "Nude" mice. Acta Pathol Microbiol Scand 1969, *77*, 758-760.

Shmulevich I, K Hunt, A El-Naggar, E Taylor, L Ramdas, P Labordé, KR Hess, R Pollock and W Zhang: Tumor specific gene expression profiles in human leiomyosarcoma: an evaluation of intratumor heterogeneity. Cancer Dev 2002, *94*, 2069-2075.

Sotiriou C, C Khanna, AA Jazaeri, D Petersen and ET Liu: Core biopsies can be used to distinguish differences in expression profiling by cDNA microarrays. J Mol Diagn 2002, *4*, 30-36.

Tapper J, E Kettunen, W El-Rifai, M Seppälä, LC Andersson and S Knuutila: Changes in gene expression during progression of ovarian carcinoma. Cancer Genet Cytogenet 2001, *128*, 1-6.

Wang GL, BH Jiang, EA Rue and GL Semenza: Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellu-lar O_2 tension. Proc Natl Acad Sci USA 1995, *92*, 5510-5514.

Wang J, TH Bø, I Jonassen, O Myklebost and E Hovig: Tumor classification and marker gene prediction by feature selection and fuzzy c-means clustering using microarray data. BMC bioinformatics 2003, *4*, 1471-2105.

Wang J, V Nygaard, B Smith-Sørensen, E Hovig and O Myklebost: MArray: analysing single, replicated or reversed microarray experiments. Bioinformatics 2002, *18*, 1139-1140.

Wolfinger RD , G Gibson, ED Wolfinger, L Bennett, H Hamadeh, P Bushel, C Afshari and R S Paules: Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 2001, *8*, 1066-5277.

Yang YH, S Dudoit, P Luu, DM Lin, V Peng, J Ngai and TP Speed: Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res 2002, *30*, 1362-4962.

Yang YH and TP Speed: Design issues for cDNA microarray experiments. Nat Rev Genet 2002, *3*, 579-588.