56th Annual Meeting of the EAAP, June 5-8, 2005, Uppsala, Sweden.
Session CG3.3, Abstract 597

Background Bias on cDNA Micro-Arrays

M.H. Pool, B. Hulsegge and L.L.G. Janss,
Division Animal Resources Development,
P.O. Box 65, 8200 AB Lelystad, The Netherlands.

Correspondence to ASGbioinformatics@wur.nl or Marco.Pool@wur.nl
Internet home page: http://www.asg.wur.nl

Abstract:
cDNA microarray experiments measure differences in mRNA expression of many genes exposed to specific conditions. Unfortunately measures are not simple reflections of differences in gene-expression but are biased by several steps in the procedure. Since microarrays are applied as a quantitative tool it should find genes with any effect. Especially for genes with smaller effects, correct prediction of deviations from the background noise becomes important. Correlations were found between foreground and local background measures ranging from 0.2 to 0.5, also across channels. To fix the foreground-local background correlations blank array spots were used, working from the assumption that the average local background should equal the average blank spot intensity. Regression of local background by color on the blank spots confirmed background bias on all of a set of 8 slides analyzed and on three levels: a constant (base line) difference, representing a systematic bias by color, a more than proportionate increase of local background vs blank spot intensities (biased by a factor 1.5 to 2), and a dependency of local background on spot foreground intensity. By fitting a local background, or preferable a surface, on the blank spots we were able to correct for the methodologically introduced background bias. This correction resulted in less negative background corrected spots, smaller error ratio's and less noise by lower spot variation.

Summary of the presentation

Introduction
cDNA Microarrays are a relative new tools and available since the late nineties in Livestock. The advantage of microarrays is that it measures thousands of genes at the same time for diseased or challenged animals. In the past, several studies have followed the candidate gene approach to find specific genes for specific traits or effects (Gerbens et al., 1998, Kramer et. al., 2003), but with micro-arrays it becomes possible to search for 'all' genes involved on one single slide – i.e. over 10.000 spots on glass slides and over 50.000 for affimetrix chips.
Besides finding 'the' genes that are involved in a certain trait or process, microarrays can also express the amount of difference in the expression of genes between animals or breeds. In this respect the usage of microarrays is not only a tool for gene discovery but is also useful in pathway analyses or in animal genetics by measuring difference in production ability due to differences in gene expressions. And also in the field of inbreeding and biodiversity microarrays might be of interest as a tool to monitor the flow of either specific or universal genes in a population.

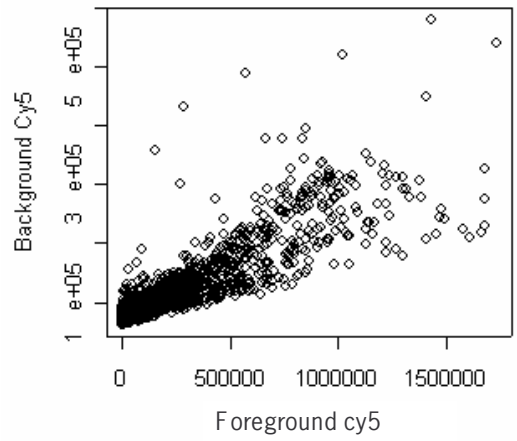


Figure 1: Local background plotted against the gene expression (foreground)

The microarray itself is constructed using (pooled) tissues from reference animals. The material printed onto the slides can be cDNA clones or are specific created oligo's. Each spot on the slide might represent a different gene but generally several spots or clones represent the same gene or parts. Above that to ensure powerful enough analysis multiple copies are printed onto the slides – normally 2 to 3 copies per slide. This means that the number of genes printed on the slides is often several times less than the number of spots. For measuring the expression of genes mRNA samples of

challenged and control animals, or any other design, are hybridized to the slides each labeled with a different fluorescence dye (Cy5 or Cy3, respectively red and green for treatment A and B).

The production of microarrays includes many steps and is therefore sensitive to several sources of variation. Good laboratory practices and correct normalization procedures are needed to certify good quality data. To facilitate the people in the lab and to standardize procedures we have developed a bioinformatics-pipeline and tested procedures for handling and normalization of microarray data (Pool et al., 2003 and Kuurman et. al, 2003). In this study we describe the quality of background correction based on local spot background and it was found that - depending on the labeling method, scanner and image analyzing software used - that the local background might be correlated with the spot intensity and therefore we suggest to include specific black spots for intensity independent background correction.

Table 1: Correlations between local background and foreground (gene expression) by different scanners, imaging software and labeling techniques.

Scanner or labelling	correlation
TSA-labeling with Analyzer 3.3	0.3 – 0.85
TSA-labeling with GenePix	up to 0.5
Aminoallyl-labeling with GenePix	up to 0.3 (50% none)

Methods and Results

For testing the presence of bias with local background correction in cDNA microarrays, slides from at least five different experiments were used (Hemert et al., 2003 Pas et.al, 2005a;2005b, Cagnazzo et al., 2005,Hulst et al). All experiments have applied microarrays printing animal tissue material on glass slide but experiments used differenced species, different labeling kits (TSA (PerkinElmer) and aminoallyl labeling), scanners (GeneTac™ 2000 (Genomic Solutions) and ScanArray Express (PerkinElmer)) or imaging software (Analyzer 3.3 (Genomic Solutions)., ScanArray Express (PerkinElmer) and GenePix (Axon Instruments)).

Since data of microarrays might be affected by not optimal hybridization or other laboratory conditions, the quality of data differs between experiments, people and over time, despite good laboratory practice and standardized procedures. One source of variation is caused by background noise for which one might need to correct. Most imaging software packages provide background corrections based on the local background spot intensity since background might vary all over the slide. By plotting and calculating the correlation between the local background against the spot intensity (foreground) it was shown that local background was biased (Figure 1, result from analyzer 3.3, and Table 1). In general there was considerable bias in local background except with the aminoallyl labeling technique correlations were lower, however high correlations were still observed in some of the arrays analyzed. To ensure that background measures will be independent of the spot intensity and hybridization conditions we introduced empty or the so called 'blank' spots. Those spots do not contain any cDNA material and would not hybridize with the mRNA sample other than some aspecific binding. Based on the blank spots the background can be expressed independent of the gene expressions and by using splines the black spot background was calculated for each spot (Pool et al, 2003).

Figure 2: Statistical analysis

Fitted models:

$$LB_R = \mu + b_1.BSB_R + b_2.FG_R + b_3.BSB_G + b_4.FG_G + e$$

$$LB_G = \mu + b_1.BSB_R + b_2.FG_R + b_3.BSB_G + b_4.FG_G + e$$

Expected:

$$LB_R = 1.BSB_R + e$$

$$LB_G = 1.BSB_G + e$$

Bias = difference

In case there is no bias the regression of the local background (LB) on foreground gene expression (FG) and the black spot background (BSB) is expected to be one (figure 2, R: red or Cy3 and G: green or Cy5). The model in Figure 2 was applied for regression on one color (results not shown) as well as regression on both colors (Table 2) and did not differ. Results (Table 2) show that the Local background did not increase 1:1 with Blank Spot Background and that it actually increased with foreground (of both colors). Therefore the local background measured near the location of the spot on the slide was found dependent on the intensity of the gene expression of the spot itself. The aminoallyl labeling – which was most recently optimized for animal tissues cDNA microarrays - showed better results, however for a number of slides correlations were increased and regression factors were significantly different from one, although the slides were indicated as a correct slides after considering the standard procedures.

Not correcting for the local background bias will lead to larger errors on estimated ratio (gene expression), additional loss of spots as the effect of higher average background correction based on local background (i.e. negative constant experiment 3, Table 2) and it creates extra noise due to larger variation (not shown).

Therefore we suggest to use Blank Spot Background (BSB) when correcting for background or no background correction at all. However, although the slide might seems to be a correct slides after all standardization procedures in the bioinformatics pipeline provided, there might be considerable correlation between local background and spot intensity. If those slides are not discarded correction based on blank spots is suggested. The next step in improving the current normalization procedure would be by including all corrections and standardization methods in 1 overall regression model.

Table 2: Regressions of local background (LB) on blank spot background (BSB) and foreground gene expression (FG) for both colors

Slide	μ	Red BSB	Red FG	Green BSB	Green FG
TSA-kit, GenePix (Red background)					
Exp 3, slide 118	-130	1.64	0.43	0.25	2.50
Exp 3, slide 126	-289	1.76	1.05	-0.38	1.38
Exp 3, slide 151	-66	1.63	0.97	0.00	0.86
Exp 3, slide 161	-90	2.30	0.96	-0.28	0.55
Exp 3, slide 163	-27	1.85	0.82	-0.11	0.89
Exp 3, slide 171	-147	1.52	-0.38	0.38	1.98
Exp 3, slide 172	-50	2.02	0.42	-0.11	1.13
Aminoallyl labeling, GenePix (Red background)					
Exp 1, slide 1	-61	0.89	0.83	0.85	-0.10
Exp 1, slide 2	58	0.27	0.81	0.01	0.06
Exp 1, slide 3	-11	0.90	0.40	0.45	0.02
Exp 1, slide 4	-62	1.05	0.37	0.65	0.23
Aminoallyl labeling, GenePix (Green background)					
Exp 1, slide 1	-102	0.13	1.10	2.14	0.06
Exp 1, slide 2	54	0.63	0.96	0.01	0.23
Exp 1, slide 3	-81	0.68	0.16	1.09	0.19
Exp 1, slide 4	-81	0.09	-0.32	1.81	0.67
Aminoallyl labeling, GenePix (Red background)					
Exp 2, slide 1	-11	1.13	0.06	0.04	0.07
Exp 2, slide 2	18	0.64	0.16	0.10	0.01
Exp 2, slide 3	-70	1.87	0.40	0.11	-0.14
Exp 2, slide 4	53	1.34	1.4	-1.04	-1.09
Aminoallyl labeling, GenePix (Green background)					
Exp 2, slide 1	-18	-0.26	-0.16	1.49	0.25
Exp 2, slide 2	11	-0.36	0.03	1.18	0.15
Exp 2, slide 3	-75	-0.45	0.06	2.43	0.14
Exp 2, slide 4	24	0.23	0.33	0.47	-0.07

References

Gerbens F., Jansen A., Van Erp A.J.M., Harders F., Meuwissen T.H.E., Rettenberger G., Veerkamp J.H., Te Pas M.F.W. 1998. The adipocyte fatty acid-binding protein locus: characterization and association with intramuscular fat content in pigs. *Mammalian Genome* **9**: 1022-1026.

Kramer J, Malek M, Lamont SJ. 2003. Association of twelve candidate gene polymorphisms and response to challenge with Salmonella enteritidis in poultry. Anim Genet. 2003 Oct;34(5):339-48.

Kuurman, W.W., M.H. Pool, B. Engel, W.G. Buist and L.L.G. Janss, 2003. A method for the analysis of cDNA microarray data. Proceedings of the 54th Annual meeting of the European Association for Animal Production, Roma, Italy, August 31st - September 3rd, session G3.3

Pool, M.H., W.W. Kuurman, B. Hulsegge, L.L.G. Janss, J.M.J. Rebel and S. van Hemert, 2003. Procedure for standardisation and normalisation of cDNA microarrays. Proceedings of the 54th Annual meeting of the European Association for Animal Production, Roma, Italy, August 31st - September 3rd, session G3.34

Pas, M.F.W te, M. Cagnazzo, A.A.C. de Wit, J. Priem, M. Pool, and R. Davoli. 2005a. Muscle transcriptomes of Duroc and Pietrain pig breeds during prenatal formation of skeletal muscle tissue using microarray technology. Arch. Anim. Breed. 48 (special issue) 141-147.

Pas, M.F.W. te ,A. de Wit, J. Priem, M. Cagnazzo, R. Davoli, V. Russo, M.H. Pool. 2005b. Prenatal muscle transcriptome expression profiling in pigs related to myogenesis. in press Journal of Muscle Research and Cell Motility

Cagnazzo, M., M.F.W. te Pas, J. Priem, A. de Wit, M.H. Pool, R. Davoli, V. Russo. 2005 Comparison of prenatal muscle tissue expression profiles of two pig breeds differing in muscle characteristics. submitted. JAS

Hemert S van, Ebbelaar BH, Smits MA, Rebel JM. Generation of EST and microarray resources for functional genomic studies on chicken intestinal health. 2003. Anim Biotechnol. Nov;14(2):133-43.

Hulst et al (Micrroarray towards guthhealth in pigs, manuscript in preparation).

ASG-Bioinformatics pipeline

Asgbioinformatics@wur.nl

Tool: Significant Analysis of Micro-arrays using Genstat

The SAMuG-tool applies the SAM procedure for detecting differentially expressed genes and indicates the proportion of falsely detected genes in a micro-array experiment. Output:

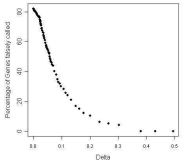
(1) List of significant genes:

```
Inputfile : C:/test.SAMmean.all.txt    Seed : 1234567

      Delta  No of pos.
            Significant          Gene Name     Score Numerator Denominator   q-value
                  Genes                           (d)       (r)     (s+s0)       (%)
                                  pl 12 E03   1.21221   3.18172   2.62473     4.178
                                   pl 2 H09   1.02294   2.66193   2.60225     4.178
                                   pl 7 C11   1.01603   2.63928   2.59763     4.178
                                   ........   .......   .......   .......     .....
                                   ........   .......   .......   .......     .....
0.302403143          90
                                   pl 8 A06   0.58180   1.60707   2.76225     5.583
                                   ........   .......   .......   .......     .....
```

(2) Delta table:

Indicating the relation between number of putative differentially expressed genes, number of significant genes, and the number of false positive estimated genes (False Discovery Rate)



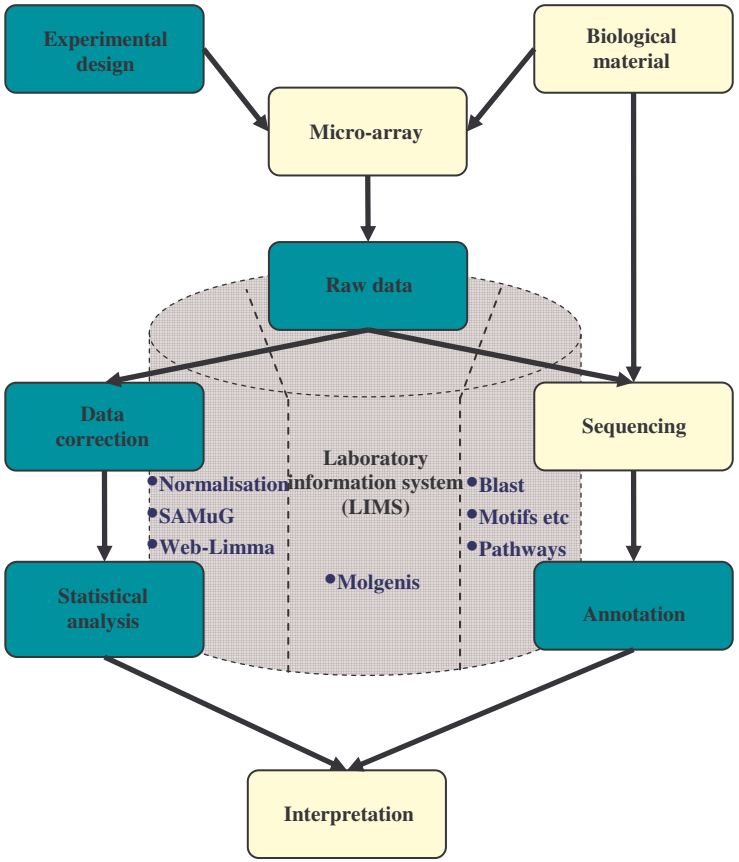
The number of significant genes is determined by what the user decides to be an acceptable False Discovery Rate by fixing the delta value. SAMuG can be used for both: single and multiple class testing.

Tool: Web_Limma

Web_Limma runs the statistical method Limma, which identifies differentially expressed genes in complex micro-array experiments. Each gene is analyzed using a linear regression model accounting for the number and variance between spots.

name	M_mutant	P.Value_mutant	Rank_mutant	M_day4	P.Value_day4	Rank_day4	M_mutant.day4	P.Value_mutant.day4	Rank_mutant.day4
pl 9 D02	-1.8127	0	2	-0.6847	0.0443	231	-0.6112	0.1204	485
pl 5 E03	-1.678	0	1	-1.0138	0.0004	28	-0.9194	0.0046	63
pl 2 E04	-1.4323	0.0003	6	-0.151	0.7201	2147	-0.2326	0.6397	1770
pl 4 D04	-1.3808	0.0011	11	0.122	0.8032	2397	-0.1294	0.8397	2414
pl 12 H03	-1.3706	0.0015	13	0.0462	0.9295	2819	-0.8316	0.0625	341
pl 16 F04	-1.323	0.0015	14	0.0917	0.8544	2575	-0.7193	0.0976	451
pl 5 E04	-1.323	0.0248	107	-0.3745	0.5503	1683	-0.9628	0.1463	539
pl 3 E06	-1.3107	0.0191	84	-0.0733	0.9189	2774	-0.081	0.9318	2798
pl 8 G02	-1.307	0.0008	8	0.1543	0.7162	2133	-0.7969	0.0466	280
pl 3 H10	-1.2864	0.0117	53	-0.4064	0.4266	1339	-0.8604	0.1211	486
pl 9 D03	-1.2661	0.0036	21	-0.1589	0.7501	2222	-0.8864	0.0493	296
pl 16 G03	-1.2652	0.0008	10	0.0287	0.9525	2873	-0.7036	0.0709	366
pl 16 D02	-1.2611	0.01	45	0.1169	0.845	2537	-0.4532	0.4543	1223
pl 5 F04	-1.1871	0.0086	40	1.1106	0.0077	98	0.0407	0.9602	2959
pl 14 A03	-1.1698	0.0268	116	0.0253	0.967	2927	-0.6133	0.3407	963
pl 5 C02	-1.1655	0.0037	23	-0.2451	0.5515	1688	0.8553	0.0429	260
pl 3 F09	-1.1606	0.0071	31	0.0913	0.863	2601	-0.3285	0.5493	1464
pl 11 G01	-1.1548	0.0008	9	-0.8365	0.0069	91	-0.5383	0.1422	526
pl 8 B02	-1.1407	0.0114	50	0.1078	0.845	2540	-0.6303	0.2168	681

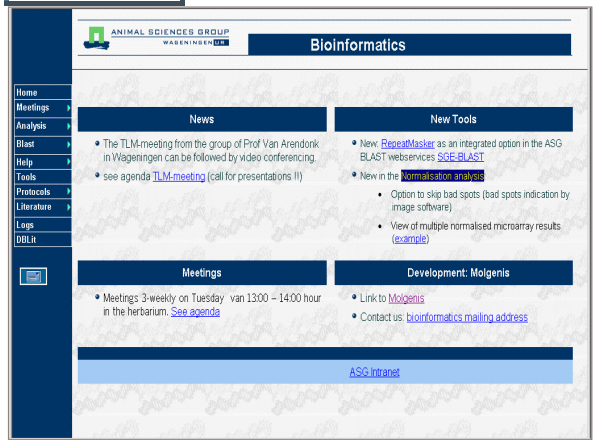
Gene effects are presented compared to the chosen reference. For each effect (main effects and interactions) a list of p-values is obtained. Per gene N p-values are given, where N is the number of effects in the model. Genes with small p-values are likely to be differential expressed, compared to the reference. Besides multiple testing Limma can also handle time-series analyses.



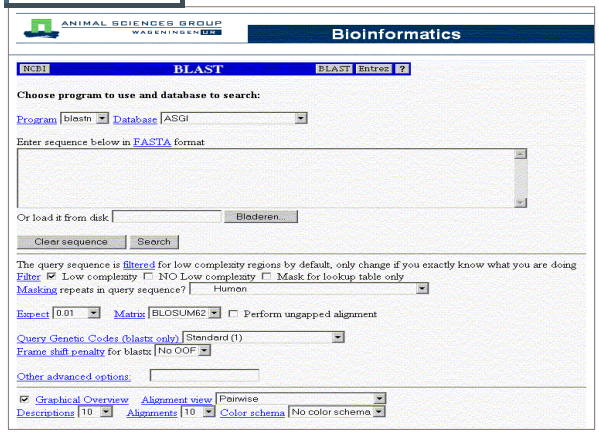
References:

- Pool, M. H., et al, 2003. Procedure for standardisation and normalisation of cDNA arrays. 54th EAAP meeting in Rome, Italy.
- W.W. Kuurman*, et al, 2003 A method for the analysis of cDNA microarray data EAAP, Rome. Italy. Abstract no. 9, Commision on Animal Genetics.
- LIMMA: Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3, No. 1, Article 3.
- SAM: Tusher et al, 2001: Significance analysis of microarrays applied to the ionizing radiation response PNAS 2001 98(9): 5116-5121,

Web Interface



Blast



Molgenis